PHYSIOLOGICAL REVIEWS

VOL. 26 JANUARY, 1946 No. 1

HISTOCHEMICAL CONTRIBUTIONS TO PHYSIOLOGY

EDWARD W. DEMPSEY AND GEORGE B. WISLOCKI

Department of Anatomy, Harvard Medical School, Boston

Histochemistry deals with the chemical characterization of substances in their natural locations within cells and tissues. As Gersh (1941) has pointed out in an admirable review of the subject, histochemistry differs from conventional biochemistry in its emphasis upon the morphological localization of the various components which collectively comprise the body substance.

Histophysiology relates the appearance of tissues, either in living states or in fixed and stained preparations, with different phases of functional activity. The principal aim of histophysiology, as opposed to physiology in general, is to provide a morphological basis for the understanding of functional events.

To follow this line of reasoning one step further, physiological histochemistry, the discipline resulting from the fusion of histochemistry and histophysiology, has as its chief aim the morphological localization within tissues of the chemical components which are involved in physiological processes and the correlation of any changes in this chemical machinery with the different functional states of the tissues or organism. In other words, it concerns itself with the ultimate description of physiological, biochemical and histological phenomena each in terms of the others. In its broader aspects, therefore, physiological histochemistry derives value not only from physical and chemical procedures, but also from the understanding of histological staining reactions in chemical terms, since variations in these staining reactions have been or may be related in turn to physiological events. In a similar fashion, histochemical observations correlated with functional states may serve to illuminate morphological problems.

In the present review we have brought together the diverse histophysiological, histochemical, and histological information available for a few situations. We have made no attempt at exhaustive coverage, but have purposely limited ourselves to a small series of topics for which numerous data derived from different fields are available for correlation. We hope, by the presentation of selected examples, to illustrate how the application of histochemical and cytochemical procedures can provide data of importance for understanding the physiological process underlying cellular activity. We hope, also, to illustrate how the various aspects of such a correlated study reinforce and validate each other so that by the combined weight of the various lines of evidence conclusions may be drawn when a single approach would prove to be indecisive.

GENERAL HISTOLOGICAL STAINING METHODS. In recent years emphasis in histochemical investigations has shifted toward the development of highly specific reactions for individual compounds or classes of substances. This trend is expressed in Lison's excellent monograph, *Histochemie animale* (1936), in which the statement is made repeatedly that certain reactions are valueless

to histochemistry because of their lack of specificity. Although, other things being equal, it is certainly true that highly specific reactions are preferable to non-specific ones, the present reviewers would like to affirm their belief that relatively nonspecific histochemical reactions which nevertheless reveal a degree of morphological localization are often of considerable value from a chemical point of view. It may be remarked parenthetically that chemists do not insist upon such extreme specificity. For example, flow birefringence, a phenomenon which is certainly not specific for any chemical grouping since it depends rather upon the axial asymmetry of the molecule, has nevertheless been used effectively in physical chemistry. A similar lack of specificity attaches to electrophoretic mobility, yet the present popularity of the Tiselius cell attests to the usefulness of the method for the solution of certain purely chemical problems. It seems axiomatic that any method which faithfully produces differential results in different regions of tissues or cells does so because of chemical or physical differences in those regions. These differences which are then responsible for the usual staining reactions of various dyes must ultimately be explained in histochemical terms. Indeed, to ignore this precept is to neglect a most important tool. It has been pointed out above that the principal emphasis in histochemistry should be upon the chemical characterization of morphological elements. The classical histological methods, on the other hand, have been devised in the main to produce the sharpest possible differential results with reference to morphological and tinctorial appearances. Obviously, therefore, the reconciliation of the specific, but frequently morphologically indistinct, reactions of histochemistry with the spatially precise, yet semi-specific, results of histology strengthens and sharpens the conclusions which can be drawn from both procedures. Moreover, from a practical standpoint, the greatest promise exists for a rapid expansion of histochemistry because there is already a vast accumulated literature dealing with the careful correlation of functional states with histological appearances. An interpretation of the conventional staining procedures in chemical terms would therefore permit a reinterpretation and extension of all of the histophysiological results presently available.

Desirable as such an interpretation of staining methods in chemical terms may be, the path to that end is fraught with many difficulties. From a chemical point of view the procedures which have been developed by histologists and cytologists, involving as they do fixation, dehydration, imbedding, clearing, mordanting, staining and differentiation, are so complex and frequently so contradictory as almost to defy chemical analysis. Indeed, it is precisely this complexity that has caused many histochemists to seek chemically simpler, though morphologically less satisfactory, methods. Nevertheless, difficult as histochemical insight may be, it is not necessarily impossible to attain. Recent advances in protein chemistry on the one hand and histochemistry on the other, as we shall attempt to unfold, offer promise of progress in this difficult field.

A great majority of histological procedures involve staining with a basic dye and counterstaining with an acid dye. The significance of the observed attachment of these two dyes to the various components of tissues has long been

obscured by a controversy as to whether the dye-substrate combination involved chemical forces or physical bonds. On the one hand, it has repeatedly been demonstrated that acidic and basic dyes react stoichiometrically with basic and acidic substances respectively. On the other hand, nonionizable substances frequently stain intensely, particularly with basic dyes. However, the solution to this particular controversy now appears to involve only the meaning of the particular terms. Colloidal chemistry is apparently governed by forces similar to those involved in general chemistry but modified into a special case as they operate on surfaces. The sign and magnitude of the surface charge differ in different physical-chemical conditions (Dubos, 1945). The charge determines the staining reaction, no matter whether the charge is due to the sum of ionic dissociations, to adsorbed ions, or to mixtures of these and other physical forces. Thus, if we understand that it is charge rather than valence which we are measuring, the reaction of dyes with tissues assumes a physical-chemical meaning. Acidophilia and basophilia, long recognized by histologists, thus appear to be designations capable of exact physical-chemical interpretation.

The acidophilia or basophilia of tissues and cells is, however, not an invariant quality. On the contrary, given structures may be selectively stained by either acid or basic dyes, depending upon the previous treatment of the tissues. Two factors in particular profoundly modify the staining properties of histological materials. These are the kind of fixation employed and the acidity of the medium in which the tissues are stained.

Among the fixatives commonly employed in histology, formaldehyde occupies a prominent place. Moreover, it has long been recognized that formaldehyde fixation renders proteins more acid and therefore increases their affinity for the basic dyes. This change in basophilia is induced by the ability of formaldehyde to combine with free amino groups, thus preventing their dissociation and increasing the effective charge caused by the dissociated carboxyl groups of the protein. On the other hand, heavy metal cations such as Hg^{++}, Al^{+++}, Cu^{++}, etc. combine with the carboxyl groups of amino acids (Tolstoouhov, 1928; Dubos, 1945) and therefore increase the acidophilia of proteins.

The effect of acidity and alkalinity upon staining reactions apparently can be accounted for satisfactorily by a consideration of ionic interchanges. McCalla (1940, 1941a, 1941b) and McCalla and Clark (1941), working with bacterial cells, have demonstrated that adsorbed cations are displaced during staining by basic dyes. If the experimental situation is arranged so that the adsorbed cations are hydrogen ions, the displacement of these ions by staining may readily be measured as a change in pH. Acid dyes have a similar capability to displace anions from the absorption shell.

The theory of ionic interchange also helps to explain some of the otherwise puzzling salt effects in staining. McCalla (1941b) has shown that Na^{+} and K^{+} displace very little hydrogen, whereas ions such as Pb^{+++}, Ag^{+}, Hg^{++} and Al^{+++} displace a great deal. The data submitted by McCalla also show that the activity of the heavy metal cation is not appreciably influenced by the kind of anion present in the system.

These observations dealing with the physical chemical aspects of the staining reactions, although fragmentary and incomplete, are exciting in their implications. From the standpoint of the conventional histological procedures, the results obtained by treatment with formaldehyde on the one hand and polyvalent heavy metals on the other are particularly significant. The shift in the isoelectric zone of many proteins which is induced by these reagents apparently causes, at usual staining pH, basophilic or acidophilic staining depending on the reagent used. Because of the affinity of formaldehyde for amino groups and of ions such as Hg^{++} for carboxyl groups, we may assume that the proteins which show such lability in their staining reactions do so because their charges are predominantly determined by amino and carboxyl groups. A corollary to this statement is that proteins the staining of which is relatively independent of the mode of fixation must depend upon other forces or more strongly dissociating groups for the determination of their charges. We shall attempt to show, in the following sections, that such strongly dissociating groups do indeed characterize at least two groups of basophilic substances, the nucleoproteins and the mucoproteins.

A. *The nucleoproteins.* The nucleoproteins are large molecules which contain a protein conjugated with nucleic acid. The nucleic acid part of the nucleoprotein apparently determines its basophilic staining reaction, since after hydrolysis the resulting protein exhibits acidophilic properties. Nucleic acids are in themselves complex substances composed of nucleotides which can in turn be broken down into component parts consisting of a purine or pyrimidine base, a sugar and phosphoric acid. Two general classes of nucleic acids exist in cells, depending upon the kind of sugar incorporated into their structure. One group, originally prepared from thymus tissue, contains a desoxyribose sugar and is frequently called desoxyribose nucleic or desoxyribonucleic acid. The second type, originally prepared from yeast cells, contains a ribose sugar and is referred to as ribose nucleic acid or ribonucleic acid. Nucleic acids are strongly charged compounds and have a great affinity for basic dyes. Moreover, their basophilia is not destroyed by conjugation with proteins, so that the nucleoproteins are also strongly basophilic. Excellent reviews dealing with the chemistry and occurrence of the nucleoproteins have recently been published by Mirsky (1943) and Greenstein (1944).

Although nucleoproteins have long been known to display pronounced basophilia and therefore to be among the substances which react with basic dyes in tissues, the positive identification of these compounds and their accurate localization in cells have only recently been accomplished. Three methods have been developed which are of the utmost utility in this regard. These are the staining method of Feulgen and Rossenbeck (1924) for desoxyribonucleoprotein, the use of ribonuclease, developed by Dubos (1937) and Brachet (1940) and subsequently improved after the enzyme had been prepared in crystalline form by Kunitz (1940), and the utilization of the specific ultraviolet absorption of nucleoproteins by Caspersson (1940).

The Feulgen method consists of the application of the Schiff test for alde-

hydes after a partial hydrolysis of the nucleoprotein in warm acid. Such gentle acid hydrolysis frees the aldehyde group of the desoxyribose sugar which then is identified by the Schiff reagent. The test, if carefully applied according to the directions of Feulgen, is apparently specific for desoxyribonucleoprotein since to date no other compound has been shown to give the reaction and the sites at which positive reactions have been obtained are in all cases compatible with the view that the reactive substance is nucleoprotein in character (cf. Stowell, 1945). Indeed, the reactions obtained by this method are confined entirely to the nucleus in animal and plant cells and to discrete nucleus-like bodies in bacterial cells (Dubos, 1945). Moreover, not all of the basophilic substances of nuclei are stained. The nucleolus, as well as certain regions of the chromosomes which stain intensely with basic dyes and which have collectively been called "heterochromatin," is prominently unstained by the Feulgen procedure.

The discovery and preparation of an enzyme capable of depolymerizing ribonucleic acid into its constituent nucleotides paved the way for the cytological localization of ribonucleoproteins. Dubos (1937) found that the basophilia of bacteria, excepting the discrete nucleus-like bodies mentioned above, was abolished after incubation in a solution containing ribonuclease, and, following Kunitz' (1940) isolation and crystallization of the enzyme, repeated the observations with the crystalline substance. Brachet (1940) described the ability of a crude preparation of ribonuclease to abolish the characteristic cytoplasmic basophilia in a variety of animal cells, including those from the pancreas, intestinal epithelium, liver and nerve cells. Desclin (1940) reported that the basophils of the anterior pituitary gland lose their affinity for basic dyes after incubation in a solution containing ribonuclease. Gersh and Bodian (1943) identified ribonucleoprotein in the Nissl bodies of nerve cells from the anterior horn of the spinal cord by the observation that these bodies lost their characteristic basophilia after digestion in a solution of crystalline ribonuclease. Also using the crystalline enzyme, Dempsey and Wislocki (1945) destroyed the basophilic reaction of the syncytium and cytotrophoblast of the human placenta and of the basophilic cells in the anterior pituitary gland of the rat. Davidson and Waymouth (1944) have recently identified basophilic inclusions in hepatic cells by means of digestion with crystalline ribonuclease.

In addition to the cytoplasmic structures listed above, a number of investigators including Brachet (1940), Gersh and Bodian (1943) and Dempsey and Wislocki (1944) have noted the disappearance of basophilic staining in the nucleolus after treatment with ribonuclease.

The development of ultraviolet microscopy and spectrophotometry by Caspersson and his collaborators has aided greatly our understanding of the localization of nucleoproteins. These compounds exhibit strong absorption at about 2600 Å, due to the presence in the molecule of purine and pyrimidine components which also show strong absorption at 2600 Å. However, although the intensity of absorption at this wave length may be used as a quantitative measure of the amount of these substances in fixed or living cells (Caspersson, 1940), the method does not discriminate between ribonucleoproteins and desoxyribonucleoproteins

since the absorption spectra of these two compounds are nearly identical. Nevertheless, the fact that the ultraviolet method provides quantitative data is of the utmost importance, and, since qualitative discrimination between ribonucleoproteins and desoxyribonucleoproteins can be made by recourse to the Feulgen method and to treatment with ribonuclease, it is now possible to obtain reliable information on the changes in concentrations of both varieties of nucleoproteins at different functional stages. Using these methods, Caspersson (1939) and Caspersson and Schultz (1940) have studied the ebb and flow of the nuclear nucleoproteins during the mitotic cycle, and Gersh and Bodian (1943) have observed the cycle of ribonucleoprotein disappearance and reappearance during retrograde chromatolysis and regeneration in the anterior horn cells of the spinal cord.

The data presented above indicate that of all the basophilic inclusions in cells from the various tissues of the body, a considerable number owe their affinity for basic dyes to their content of nucleoproteins. Moreover, it is now possible to distinguish between two types of nucleoprotein, the desoxyribonucleoproteins, which apparently are confined entirely to the nuclear chromatin, and the ribonucleoproteins which occur as characteristic cytoplasmic inclusions in many cells as well as in the nucleolus and other nuclear locations. We shall attempt to show in the following sections that another large group of compounds, the mucoproteins, also exhibit basophilia and may be identified in cells by their cytological location and by their characteristic reactions.

B. *The mucoproteins.* The mucoproteins are strongly acid, complex substances which on hydrolysis yield a protein, a hexosamine-carbohydrate and, as a rule, sulfuric acid. Those which contain sulfuric acid are therefore to be regarded as sulfuric esters of glycoproteins and are strongly basophilic in their staining reactions. They are either soluble or exhibit pronounced swelling in water, and they are easily hydrolyzed to free their constituent carbohydrate molecules. These properties have been used with considerable success in devising methods for their identification.

The strong acidity of mucus, apparently caused by the presence of sulfuric acid, accounts for the marked basophilic properties of the compounds. Indeed, the basophilia exhibited by these compounds may be revealed even after staining in dye solutions so dilute that nuclear chromatin is unstained (Hempelmann, 1940). Although the mucoproteins have a great affinity for basic dyes, their basophilia exhibits a peculiar quality in that the tint assumed by the stained structure may be different from that of the free stain. This property, usually referred to as metachromasia, is especially noteworthy with some of the thiazin dyes such as toluidin blue and polychrome methylene blue. The phenomenon has been extensively studied from a histochemical viewpoint by Lison (1935) who regards metachromasia as indicative of a sulfuric ester of high molecular weight, since in no other class of compounds could the effect be observed. Michaelis (1944) has adduced evidence that the phenomenon is caused by a mechanism whereby the dye is attached in a peculiar spatial pattern.

Hempelmann (1940) has recently published the results of an excellent study

of the staining reactions of the mucoproteins. To determine the specificity of the metachromatic reaction after staining with toluidin blue or polychrome methylene blue, he stained numerous sections of a wide variety of tissues with dilute solutions of these dyes. The tissues in which the red-violet metachromatic reaction was observed were *1*, the intracellular matrix of cartilage; *2*, the walls of arteries that had undergone mucoid degeneration; *3*, the ground-substance of heart valves and some of the fibers of the atrioventricular ring; *4*, the sclera of the eye; *5*, tendon; *6*, intracellular mucus and the secretion of the epithelial mucous glands; *7*, Wharton's jelly in the umbilical cord, and *8*, the mucoid substance in egg yolk. To this list should be added the granules of mast cells and basophilic leucocytes (Holmgren and Wilander, 1940), certain decidual cells in the maternal placenta (Asplund, Borell and Holmgren, 1940) and the mesenchymal stroma of embryos (Holmgren, 1940). This list includes practically all of the tissues from which mucoproteins have been extracted chemically.

In addition to the selectivity for mucoproteins which is displayed by the metachromatic reaction, Hempelmann found that two varieties could be differentiated on the basis of their affinities for dilute solutions of the two dyes. When a solution of either dye diluted to 1:250,000 was used, the glandular mucus as well as the mucoprotein associated with the connective tissues was stained vividly. When greater dilutions (up to 1:1,280,000) were employed, the epithelial mucus failed eventually to take the stain while the mesenchymal mucoproteins stained almost as intensely as before. These two types of mucus differ in that epithelial mucus contains mucoitin sulfuric acid and a protein, whereas mesenchymal mucus is a complex containing chondroitin sulfuric acid. In order to explain the differential staining at great dilution, it is necessary to assume that the chondroitin sulfuric protein has a greater affinity for the dye than does the mucoitin complex.

The ready solubility of mucus in water prevents its preservation in many histological fixatives. Mixtures containing alcohol are consequently recommended. For the best demonstration of the metachromasia of mucous substances, fixatives containing the salts of heavy metals are advantageous. Jorpes, Holmgren and Wilander (1937), Holmgren and Wilander (1940) and Holmgren (1940), who have made extensive studies on the metachromatic granules of mast cells and on the occurrence and distribution of metachromasia in the tissues of adults and embryos, have relied for the most part on basic lead acetate as a fixative.

In addition to staining metachromatically, the mucoproteins may be identified by tests which reveal the carbohydrate moiety of the molecule. The carbohydrate is easily unmasked by mild hydrolysis and may then be stained in a number of ways. Thus Schiff's reagent for aldehydes gives an intense purple reaction. Since this reagent also stains glycogen (Bauer, 1932, 1933), the mucoprotein carbohydrate moiety must be differentiated from the former either by reference to the relative solubilities of the two compounds after fixation or, better still, by the saliva test which destroys glycogen but leaves the carbohy-

drate of mucoproteins unaffected. The silver reduction test for glycogen employed by Mitchell and Wislocki (1944) also gives positive results for mucoproteins after suitable fixation. Here again glycogen and mucoproteins may be differentiated by means of the saliva test. Still another reaction which depends also upon the carbohydrate part of the mucus molecule is the test for glucosamine originally developed by Ehrlich (1901). This test involves the formation of a red color with dimethyl-paramido-benzaldehyde after mild alkaline hydrolysis. This test is superior to the tests for reducing agents listed above in that it selectively reacts with glucosamine (Mann, 1902).

The mucoproteins, therefore, represent a class of compounds which exhibit strong affinities for the basic dyes. They are frequently poorly preserved in histological sections unless specific attention is given to their fixation. If adequately fixed, they may easily be recognized by their metachromasia after staining with certain dyes. Gentle acid or alkaline hydrolysis splits them to form reducing bodies which give the characteristic reactions of glucosamine.

C. *Other basophilic inclusions.* In addition to the cytoplasmic inclusions which have been sufficiently investigated to permit their identification as nucleoproteins or mucoproteins, a variety of other basophilic structures exists the nature of which is undetermined at present. For example, the thyroid colloid shows strong basophilia, the intensity of which varies in different physiological states. Although nucleoprotein is known to exist in the colloid (Salter, 1940), it remains to be determined whether this substance accounts completely for the basophilic reaction. Finally, attention should be called to the fact that certain structures exhibit basophilia although auxiliary tests have shown neither nucleoprotein nor mucoprotein to be present. The cross-striations of skeletal muscle are noteworthy in this regard. These structures stain intensely with basic dyes after fixation in either formaldehyde or in salts of heavy metals, their staining reaction is not metachromatic, they do not exhibit a positive Feulgen reaction, and digestion with ribonuclease does not affect their affinity for basic dyes. It must be concluded from these characteristics that the substance responsible for the basophilia of the striations does not depend for its staining characteristics upon the presence of carboxyl groups, and that it is not mucoprotein, desoxyribonucleoprotein or ribonucleoprotein.

D. *General considerations relating to basophilia.* The considerations presented in the preceding sections point to the possibility that the generalized histological phenomenon of basophilia may be broken down into several categories depending upon specific histochemical reactions. Moreover, the simultaneous comparison of the histological and histochemical preparations provides precision of localization with chemical specificity.

It seems reasonable to divide the basophilic proteins into two main classes. The first class consists of proteins, the isoelectric point of which is shifted to the alkaline side of the ordinary staining pH by fixation with the salts of heavy metals and to the acid side by fixation in formaldehyde. The known combining powers of these fixatives make the assumption reasonable that these changes in isoelectric points are caused for the most part by rendering the carboxyl and amino groups, respectively, non-dissociable.

The second class consists of proteins in which the isoelectric point is more acid than the staining pH after reaction with the salts of heavy metals. Such behavior indicates the presence in the molecule of a strongly acid group, and in the case of two groups of proteins which fall within this class, strongly acid groups are known to be present. Thus the nucleoproteins are phosphoric esters and the mucoproteins are sulfuric esters. It should, however, be pointed out that the present information does not preclude the presence of other basophilic groups in these molecules. For example, it is quite likely that fixation by heavy metals shifts the isoelectric point of nucleoproteins in the alkaline direction, albeit not so far as to reach the pH at which histological staining reactions are ordinarily carried out.

These considerations point a direction for two future lines of research. If valid, they indicate that strongly acid radicals should characterize the still poorly understood basophilic compounds such as those responsible for the basic staining of the cross-striations of skeletal muscle. In this connection, it is well known that a variety of phosphate-containing substances such as phosphocreatine and adenosine triphosphate is present in muscle and are importantly concerned with the chemistry of muscle contraction. The second, more general, line of research involves the characterization of histologically demonstrable proteins by determining their isoelectric points and by the displacement of these isoelectric points by treatment with reagents such as formaldehyde, heavy metal ions, etc. Hence, determining the pH at which a given cellular inclusion changes from basophilic to acidophilic staining should provide data for comparison with similar information derived from the study of purified compounds in vitro.

Reactions of lipoidal materials. Both the histochemical localization and identification of fat-soluble substances are of considerable importance for the study of cellular physiological processes. The absorption, transport, storage and metabolism of fats all pose questions involving accurate localization and identification. Moreover, a number of lipoidal compounds of biological importance are now known (e.g., the steroid hormones and the fat-soluble vitamins). Finally, modern views of cellular organization envision the presence of protoplasmic interfaces on which surface reactions may occur and lipoids are apparently of great importance in the production of such interfacial surfaces.

Among the general methods for staining lipoids, two procedures have been quite popular. These are the reduction methods in which silver or osmic acid is reduced to a black precipitate, and the solubility methods in which the fatty materials are colored by a fat-soluble but water-insoluble dye. The reduction methods are capricious and non-specific, but, on the other hand, are exquisite in the morphological detail which they reveal. The solubility methods which usually employ Sudan III or Sudan IV produce somewhat less precise localization but provide information on the solubility characteristics of the structures which are stained. Moreover, in recent years the use of a new dye, Sudan black B, has overcome some of the limitations of the earlier procedures. It has long been known that the lower members of the Sudan series do not stain all lipoids, since they are relatively insoluble in such lipoidal substances as the phospholipids and cholesterol. In addition, the color of the red Sudan dyes is not suffi-

ciently intense to permit resolution of the smallest lipoidal inclusions. Sudan black, however, generates an intense stain in myelin, in the Golgi apparatus, and in other lipoidal inclusions which are not revealed by the lower members of the Sudan series (Baker, 1944).

Accurate chemical characterization of the lipoidal materials in microscopic sections is not yet possible. Nevertheless, a beginning at least has been made and several group reactions are now known. The Smith-Dietrich test for lipines (Dietrich, 1910) is apparently one of these (Lison, 1936). Although the test is not absolutely specific in a chemical sense, compounds such as lecithin, cephalin and sphingomyelin resist the differentiation imposed in the test to such a degree that a positive reaction is very strong presumptive evidence for their presence. Baker (1944) has recently adduced evidence from this and other reactions that a nitrogenous lipoid is present in the Golgi apparatus of several diverse cells.

One of the reactions which characterizes true fats, as well as other lipoidal substances, is saponification. This reaction, which results in the formation of soaps, does not appear to have received the attention it deserves. In plant tissues, however, the formation of soaps by alkaline hydrolysis of fats has proved useful (McClung, 1937). Perhaps the reason for its present disrepute stems from the erroneous belief of earlier investigators that the method differentiated free fatty acids from their esters (Fischler, 1904). This it certainly does not do (cf. Lison, 1936, pp. 203–204). Nevertheless, a classification of fatty inclusions based upon the ease or impossibility of their saponification should prove useful, particularly in tissues in which both saponifiable and non-saponifiable lipoids are simultaneously present. The method should also prove useful for removing saponifiable material from sections leaving the non-saponifiable fraction for subsequent study.

The recent development of the related fields of steroid chemistry and endocrinology has prompted several investigations of the various glands which secrete the steroid hormones. These hormones are all alike in that they contain a cyclopentenophenanthrene nucleus. The various compounds which have diverse biological effects differ in the kind and location of various side groupings on the nucleus. The activity appears to be due to hydroxy- or keto-groupings, for the most part. Reasoning that ketone reactions, coupled with solubility tests, should reveal the location of the ketosteroids, Bennett (1940) found that the outer layer of the fasciculata in the adrenal cortex of the cat contained acetone-soluble materials which formed phenylhydrazones and semicarbazides. He showed also that the lipoids in this area were capable of reducing silver in alkaline solutions. The lipoidal droplets contained within the cells were frequently birefringent, and, on reaction with digitonine, anisotropic crystals were formed. Since these reactions are characteristic of ketosteroids, and since they were negative in regions of the adrenal cortex other than the outer layer of the fasciculata, Bennett concluded that this zone represents the site of origin and secretion of the adrenal cortical hormones. Subsequently, the phenylhydrazine reaction has been applied to the mammalian testis where

phenylhydrazones are formed in the interstitial cells (Pollock, 1942), to the rat's ovary, where the reacting substances are restricted to thecal, luteal and interstitial cells (Dempsey and Bassett, 1943) and to the human placenta in which reactions have been observed in the syncytial trophoblast (Wislocki and Bennett, 1943).

Bennett's conclusions were severely criticized by Gomori (1942) who pointed out that the phenylhydrazones formed after treatment with phenylhydrazine were superimposable upon the results obtained by the plasmal reaction of Feulgen and Voit (1924). The plasmal reaction involves the recolorization of leucofuchsin, which, under the name of Schiff's reagent, is commonly employed in tests for aldehydes. Feulgen and Voit (1924), Verne (1929), and Gomori (1942) regard the reaction as indicating the presence of aldehydes of fatty acids. However, it has not been possible to oxidize fatty acids by oxidations as mild as those which are effective in the plasmal reaction, and Lison (1932) in a comprehensive study found that positive reactions with Schiff's reagent could be obtained with aldehydes, certain ketones, and even by some unsaturated compounds which contained no carbonyl grouping. Moreover, the carbonyl group in certain steroids is active enough to react with phenylhydrazine (Bennett, 1940), Schiff's reagent (Dempsey and Wislocki, 1944), and silver in alkaline solution (Reichstein and Shoppee, 1943). It therefore appears that these reactions, commonly employed to detect aldehydes, are not capable of differentiating between aldehydes and ketosteroids. Yet, accessory evidence to be presented below, indicates that at least in the adrenal gland, ovary, testis and placenta, positive phenylhydrazine reactions occur in the same locations which also may exhibit other steroid reactions and are therefore indicative of steroid substances.

Among the characteristic reactions of the steroid hormones, their fluorescence (Bierry and Gouzon, 1936; Reichstein and Shoppee, 1943), their ability to form spherocrystals, their reaction with digitonine to form crystals which are anisotropic (Fieser, 1937), and their response to the Liebermann-Burchardt test are noteworthy. These procedures have been applied in this laboratory to the ovary (Dempsey and Bassett, 1943) and the placenta (Dempsey and Wislocki, 1944). In brief, positive reactions have been obtained in the same locations which also contain substances reacting with phenylhydrazine and Schiff's reagent.

Since the active hormones elaborated by the organs studied are steroids possessing keto-groups, it seems reasonable to attribute both the positive steroid reactions and the reactions with phenylhydrazine and Schiffs reagent to the hormones or their precursors. At all events, the point of doing these tests is to locate the areas of the glands which contain biologically active substances. Even if some of the positive areas contain additional substances such as aldehydes which add to the intensity of reaction, or if the adventitious aldehydes turn out to be more widely distributed than the hormones themselves, it seems that one can safely say that the non-reactive areas do not contain the hormones in significant amounts.

The correlation of the steroid reactions described above with different physi-

ological states has so far been done only fragmentarily. Sarason (1943) has reported that the sudanophilic and birefringent zone in the adrenal cortex decreases in thickness after hypophysectomy. The changes in the adrenal cortex during the alarm reaction have been studied by Popjak (1944). In moderate stress there is an increase in the phenylhydrazine reaction, whereas during severe stress a depletion of the substances which form phenylhydrazones occurs. These results have been confirmed in this laboratory (Deane, unpublished experiments) for both hypophysectomized rats and rats subjected to cold environments. The changes observed after these experimental procedures may be detected either with the Schiff reaction or by observing the birefringence or fluorescence of the gland. For the ovary of the rat, Dempsey and Bassett (1943) described the onset of the various reactions at approximately the time at which the Graafian follicle develops its antrum. They also called attention to changes in the reactions which occur during the lifetime of the corpus luteum, but made no detailed analysis of these changes in reference to the physiological cycle of the ovary. However, Everett (1945) has recently correlated the Liebermann-Burchardt reaction with the stage of the reproductive cycle in rats and has shown that a marked change in the reactiveness of the corpus luteum of the previous cycle normally occurs just prior to ovulation. In persistently estrous rats, in which ovulation does not spontaneously occur, there is an almost complete deficiency in the reactive material of the corpus luteum. Since Everett has previously shown that the persistently estrous rats fail to ovulate because of a preovulatory deficiency of progesterone, there is presumptive evidence that the Liebermann-Burchardt reaction may serve as an index of progesterone formation and secretion. This very promising lead should be investigated further. So far as the testis is concerned, relatively little is known. Talbot and Dempsey (unpublished observations) have found little or no birefringent lipoids in the interstitial cells of infants and young boys, whereas considerable amounts are present in the Leydig cells of adult men. In one case of marked sexual precocity, birefringent lipoids were observed in the testis of a child. Conversely, birefringence of the testicular lipoids was decreased or absent in several cases of hypogonadism.

Before passing from the subject of the steroid reactions, it may be appropriate to mention a valuable improvement on the phenylhydrazine hydrochloride method of Bennett which has been worked out in this laboratory by M. Pechet, but which has not yet been published. This method involves the use of 2-4 dinitrophenylhydrazine sulfate in phosphate buffer at pH 6.0–6.8. The phenylhydrazones formed with this compound are much deeper in color than are those formed after treatment with phenylhydrazine hydrochloride. In consequence, thinner sections can be prepared and much better localization of the reactive substances is possible than following the previous method. Moreover, the nitro-derivative reacts more energetically than does phenylhydrazine, a circumstance which has been utilized to advantage in identifying the less reactive carbonyl groupings in sections of the adrenal cortex by the combined use of the hydrochloride and the nitro-derivative. The procedure devised involves treatment

of the section with phenylhydrazine hydrochloride until the more reactive groups are combined, followed by subsequent immersion in a solution of 2-4 dinitrophenylhydrazine sulfate. The red dinitrophenylhydrazone may be differentiated from the yellow phenylhydrazone by virtue of the difference in their colors.

Vitamin A has recently received considerable attention because of the possibility of localizing it by virtue of its specific fluorescence. This subject has recently been reviewed by Popper (1944). However, the characteristic greenish fluorescence of vitamin A may be confused with the similar greenish fluorescence exhibited by many steroids. Nevertheless, the two may be differentiated in a variety of ways. The fluorescence of vitamin A fades rapidly on exposure to ultraviolet light, whereas the greenish fluorescence identifiable in the organs which secrete steroidal compounds is stable and remains unaltered after hours of exposure. Moreover, the Carr-Price reaction, as modified for tissue sections (cf. Cowdry, 1943), provides a color test which is specific for vitamin A. Another chemical reaction which has so far not been applied to sections, has recently been published by Sobel and Werbin (1945). This test involves the development of a blue color when vitamin A reacts with glycerol 1,3-dichlorohydrin. The technique is somewhat simpler than is that in the Carr-Price reaction, the color is more stable, and carotene may also be detected as this latter substance produces a green color.

ENZYMATIC REACTIONS AS APPLIED TO TISSUE SECTIONS. The study of enzymes and enzymatic reactions has received increasing attention from biochemists in recent years. It is now generally believed that most of the chemical reactions which occur in living cells are controlled by enzymes. The enzyme exerts a directing influence by making the reaction proceed so rapidly in one direction that side reactions become of negligible importance (Sumner and Somers, 1943). It is apparent, therefore, that studies of the concentration and localization of enzymes provide information which can be related directly to the physiological activity of the tissues. Unfortunately only very few of the numerous enzyme reactions in the armamentarium of the chemist are at present applicable as true histochemical reactions. A certain number of microchemical procedures have been devised and have been carried out with ingenious histological control by the methods of Linderstrøm-Lang (cf. Biol. Symp., 1944, Ann. Rev. Biochem., 1944 for references), but the degree of cytological localization possible with these procedures is actually very slight. Among the methods which have successfully applied to sections, those concerning a variety of oxidases and peroxidases have received the greatest amount of attention.

A. *Oxidative enzymes.* The fact that organs differ in their oxidizing powers has been a part of biochemical knowledge since Ehrlich (1885), who showed that the injection of substances now described as reduction-oxidation indicators stained some organs but not others. Among the reactions which he employed was the formation of indophenol blue upon injection of alpha-naphthol and paraphenylenediamine. This reaction is usually described in current literature as the indophenol oxidase reaction, the nadi reaction, or simply as the oxidase reaction. It has been applied to a large number of animal tissues. In recent years

this reaction has attained even greater popularity since Keilin (1933) demonstrated that it reveals the presence of Warburg's *Atmungsferment*, now ordinarily described as the cytochrome oxidase-cytochrome c system.

The indophenol oxidase reaction, although it apparently reveals the cytochrome system, must be carefully employed to give reliable results. Substances occur, notably in cells of the myeloid series, which are responsive to the reaction after drying or other types of fixation. On the other hand, more labile substances exist in the cells of many other tissues, the reaction in which can only be obtained in fresh unfixed tissues. Moreover, the pH at which the reaction is carried out markedly influences the results. Lison (1936) regards only the reaction carried out in alkaline solution on fresh tissues as indicative of the presence of the cytochrome oxidase-cytochrome c system.

In addition to the nadi reaction, similar results have been obtained by the use of other phenolic substances as substrates for the enzyme system (Lison, 1936).

The oxidase reactions, as employed above, are capable of utilizing oxygen in its atmospheric or molecular form. In contrast to the oxidase enzymes, another set of enzymes commonly known as the peroxidases, are capable of oxidizing the same substrates but only in the presence of nascent oxygen, the source of which is ordinarily hydrogen peroxide. The presence of true peroxidases was formerly doubted in mammalian tissues, since the ubiquitous occurrence of catalase together with the extreme toxicity of peroxide was thought to preclude any biological importance of the peroxidase system. However, it now seems well established that peroxidases are a normal constituent of mammalian cells (cf. Lippman, 1943) and a tentative function has been suggested for the enzyme in the metabolism of iodine in the thyroid gland (Dempsey, 1944). In general, it may be said that the oxidase and peroxidase reactions reveal the specific activity of the respective enzymes. This statement appears to be true by definition, since the test actually involves the catalytic coupling of oxygen with the various substrates. Moreover, the results obtained by these reactions are in good agreement with those secured by the use of reduction-oxidation indicator dyes (cf. Flexner and Stiehler, 1938). So far as histological localization is concerned, the reactions appear to be entirely adequate, since the colored oxidation products of the substrates are highly insoluble substances and are therefore precipitated out in the cells in which the enzyme occurs. However, there is some question as to the cytological precision of the tests. Lison (1936, p. 278 et seq.) has reviewed this problem, a short résumé of the elements of which is discussed here. In brief, the problem involves the possibility that elements exist within the cell which show an extreme affinity for the oxidation products of the substrate. Hollande (1924) has presented experiments showing that identical effects can be obtained either by the nadi reaction or by staining with an alcoholic solution of indophenol blue, diluted at the moment of staining with water. Such dilution of the dye causes it to assume a metastable condition from which it is easily precipitated. The precipitation apparently occurs selectively upon certain intracellular granules, and this gives rise to the typically

granulated appearance characteristic of the nadi reaction. These granules, according to Hollande, are, properly speaking, not oxidase granules but indophenophile granules. In the case of the peroxidase reaction in which benzidine is used as a substrate, the granules are oxybenzophile.

Regardless of the degree of cytological precision possible with these methods, they provide, nevertheless, information which may be correlated with the functional state of individual tissues. Flexner and Stiehler (1938) have observed that a marked increase in the oxidase reaction corresponds in time to the beginning of secretion by the chorioid plexus of the pig embryo. A similar situation obtains in the developing renal tubules (Flexner, 1939). Dempsey and Wislocki (1944), in studying the distribution of glycogen in the human placenta, noted that glycogen accumulation occurred in tissues the oxidase activity of which was low. They also pointed out, from comparing Weed's (1917) data on glycogen storage in the cells of the developing chorioid plexus with Flexner and Stiehler's (1938) information concerning the occurrence there of oxidase reaction and reaction-oxidation potentials, that glycogen precedes the advent of the oxidase reaction, the two processes being apparently mutually incompatible. A somewhat similar relationship exists in the developing renal tubules. These considerations and others led Dempsey and Wislocki (1944) to postulate that the significance of glycogen storage in cells of the placenta and fetal organs might be to furnish energy for anerobic respiration in cells the respiratory metabolism of which was insufficient for their needs.

Both the oxidase and peroxidase reactions have also been correlated with the activity of the thyroid gland. Blue granules can easily be demonstrated in the thyroid cells after application of the nadi reagents. The peroxidase, or benzidine-H_2O_2 reaction, also causes blue granules to appear in the thyroid cells. The latter reaction is abolished by the addition of thiouracil to the bath in which the sections are stained, whereas the oxidase reaction is unaffected by this drug. Since thiouracil prevents the formation of thyroid hormone by the thyroid gland (Astwood, Bissel, Sullivan and Tyslowitz, 1943; MacKenzie and MacKenzie, 1943) and since peroxidase catalyzes reactions analogous to those involved in thyroxin synthesis (Westerfeld and Lowe, 1942), Dempsey (1944) suggested that peroxidase was a natural component of the thyroid cell and that it participated in the reactions leading to the formation of thyroxin. On the other hand, the reduction-oxidation potentials of the thyroid gland undergo changes in different physiological states (DeRobertis and Goncalves, 1945), and poisons, such as cyanide and azide which inhibit the cytochrome system, also inhibit the synthesis of thyroxin in surviving slices of thyroid tissue (Schachner, Franklin and Chaikoff, 1943). The latter authors have suggested, therefore, that the cytochrome system provides the oxidative energy for the synthesis of the thyroid hormone.

B. *The dopa oxidase reaction.* An enzyme exists in melanin-forming cells which is capable of catalyzing the oxidation of desoxyphenylalanine. This substance, usually abbreviated to dopa, polymerizes when oxidized to form a tarry compound similar to, if not identical with, the natural pigment (Bloch,

1917). The dopa reaction as carried out on tissue sections involves the incubation of the section in a solution of dopa. The presence of dopa oxidase is indicated by the formation of a black or brown insoluble pigment at the site of the enzyme.

In recent years interest in pigment formation has been revived because of the discovery of physiological methods for the control of melanogenesis. Ralli and Graef (1943) have described changes in pigmentation induced by altering the adrenal cortical function of rats. Similarly, graying of the hair has been described as a result of deficiency in certain B vitamins (cf. Unna, Richards and Sampson, 1941). So far as the present reviewers are aware, however, systematic studies of the relation of the dopa reaction to the factors causing changes in pigmentation have not yet been made. Such studies should contribute greatly to our understanding both of the mechanism of melanogenesis and the physiological effects of the vitamins and hormones on pigmentation.

C. *Phosphatases.* The simultaneous publication in 1939 by Gomori and Takamatsu of nearly identical methods for the histological demonstration of alkaline phosphatase opened a new chapter in enzymatic histochemistry. The methods involve the incubation of tissue sections in a buffered solution containing glycerophosphate as a substrate and calcium ions. Under these conditions the enzyme liberates phosphoric acid from the phosphate. The phosphoric acid immediately reacts with the calcium ions and, in an alkaline medium, precipitates out as tricalcium phosphate. The latter substance is finally revealed by transformation into a black precipitate by treatment with cobaltous salts and ammonium sulfate (Gomori, 1941a), by staining with alizarin (Kabat and Furth, 1941), or by silver impregnation methods (Bourne, 1943).

Soon, following the publication of his method for alkaline phosphatase, Gomori (1941b) announced a similar method for the demonstration of acid phosphatase. Since most inorganic phosphates are soluble in acid solution, it was necessary to find one which was insoluble at pH 4.7, the optimum of the enzyme. Gomori found that lead phosphate was sufficiently insoluble to meet these requirements and that lead salts did not inhibit the reaction.

The discovery of the two methods for alkaline and acid phosphatase led very shortly thereafter to the publication of three valuable survey papers (Gomori, 1941a; Kabat and Furth, 1941; Wolf, Kabat and Newman, 1943) in which the principal sites of phosphatase activity in the mammalian body were described. It was immediately apparent from these studies that phosphatases were widespread in their occurrence in the body. In various locations phosphatase occurred variously within cell nuclei, in the cell cytoplasm, in tissue spaces and in the lumina of glands. However, when different species were compared one with another, marked differences in distribution and concentration of the enzymes became apparent. For example, in the liver, the reaction occurs variably and unpredictably in the endothelial cells, the hepatic cells and the bile capillaries of different species. Similarly, in the ovary of the rabbit, the granulosa cells contain alkaline phosphatase and the thecal cells none, whereas in the sow and monkey, the enzyme is restricted to the theca and does not occur in the granulosa (Corner, 1944).

Despite such apparent contradictions in different species, there is reason to believe that phosphatases are importantly concerned in various metabolic processes. Thus, in developing bone, there is an alkaline phosphatase, the purpose of which is apparently to provide phosphate ions for the calcificatory process (Robison, 1922; Harris, 1932; Horowitz, 1942). Similarly, phosphatases are believed to play rôles in the metabolism of nucleotides, phospholipids and carbohydrates (Sumner and Somers, 1943). We do not wish to discuss these relationships fully here, but will return to the possible significance of phosphatase to these metabolic processes in a later section.

Emmel (1945) has recently made the interesting observation that alkaline phosphatase is concentrated in the Golgi zone of the absorptive cells of the intestinal mucosa. The amount of the enzyme present in the Golgi apparatus of these epithelial cells is so great that the phosphatase preparations rival those prepared by osmic acid methods in the elegance with which the Golgi apparatus is displayed. We have confirmed Emmel's observation regarding alkaline phosphatase (Deane and Dempsey, unpublished data) and have made the additional observation that acid phosphatase is also prominent in the Golgi region of intestinal cells. Moreover, both acid and alkaline phosphatase preparations differ in one interesting respect from the Golgi preparations made by osmic acid or silver reduction. With these latter methods, the Golgi net is most prominently and easily impregnated in the cells covering the tips of the villi, whereas the cells lining the sides and deeper parts of the villi are more difficult to demonstrate (Weiner, 1928). Conversely, in the phosphatase preparations, the Golgi apparatus in the cells at the tips of the villi is never impregnated, whereas it is easily demonstrable in the cells at the sides and base. Emmel was unable to demonstrate alkaline phosphatase in the Golgi region of kidney epithelium and therefore regarded the phosphatase reaction in the Golgi apparatus of the intestinal epithelium as unique. Nevertheless, although more difficult to achieve, we have been able to demonstrate both acid and alkaline phosphatase in the form of granules or even occasionally as well-defined nets in the zone occupied by the Golgi apparatus in a variety of epithelia including the liver, the kidney, the pregnant uterus, the seminal epithelium and the epididymis. These observations afford a specific chemical clue to the significance of the Golgi apparatus. On purely cytological grounds the Golgi apparatus has long been regarded as a center of vital cellular and secretory activity, partly because of its invariable enlargement in states of heightened physiological activity and partly because secretory granules appear to be formed in association with it or at least in its vicinity (Kirkman and Severinghaus, 1938). The observations cited above seem to substantiate this view on cytochemical grounds, indicating as they do a close association between the Golgi apparatus and enzymes known to participate importantly in cellular metabolic processes.

An interesting example of how chemical inferences may be drawn from histological data is derived from the action of different fixatives upon phosphatase. Gomori (1941b) remarks that phosphatase cannot be demonstrated in sections fixed in standard histological fixatives such as formaldehyde or mixtures which contain such substances as mercuric chloride or picric acid. One can conclude

from these facts that for its proper action the enzyme requires both free amino and carboxyl groups, since, as we have already discussed in an earlier section, formaldehyde and the heavy metals used as fixatives have the properties of combining respectively with amino and carboxyl groups and preventing their dissociation. Consequently, from a histochemical point of view, considerable interest attaches to a recent paper by Gould (1944) in which after purely chemical procedures it was concluded that phosphatase requires the presence of free amino and carboxyl groups.

CHEMICAL INTERRELATIONSHIPS IN CELLS: COMPARATIVE CYTOCHEMISTRY. It is a well-recognized principle in histology that organs or tissues may be regarded as the sum of a relatively small number of different structural elements comprising the component cells and interstitial materials. The morphological characteristics of the cells and interstitial materials serve to differentiate different tissues one from another. For example, the principal distinction between bone and fibrous connective tissue might be expressed by the statement that the interstitial matrix of the former contains dense deposits of calcium salts whereas the latter does not. A system of classification based upon these principles was, as a matter of fact, proposed by Oliver Wendell Holmes in 1862 together with the suggestion that suitable symbols might be attached to the various elemental morphological constituents so that histological structures could be represented by formulae similar to those which have proved useful in chemistry. This suggestion never caught on, probably because the number of elemental components which could be agreed upon was so small as to make such symbolic representation unnecessary.

At the cytological level of study, some systematization based principally upon similarities of morphological appearance and reactivity to staining procedures has been achieved in the past. The classification of the cells of the organism into the component tissues of the body is such a systematization (Bichat, 1801). However, the available criteria are in most instances so coarse that only the crudest kinds of similarities are enjoyed by some of the members of the same group. For example, the epithelia are so widely divergent from one another both in structure and apparent function that relatively little is to be gained by grouping the parenchymal cells of the thyroid with, say, those of the epidermis into one category. Such discrepant aspects within the customarily recognized divisions have led to a kind of systematic empiricism shared by histologists and cytologists whereby only lip-service is paid to the category, and each organ or indeed each of several separate cell types which may occur within it is regarded as a separate and unique phenomenon.

The introduction of histochemical and cytochemical methods permits another system of classification or characterization of cells. It is possible, for example, to group into one category the cells in which basophilic cytoplasmic nucleoproteins are an outstanding characteristic, into another category those cells which contain large amounts of phosphatase, and into a third group those which exhibit glycogen storage. This system is capable of indefinite expansion as other

methods reveal further histochemical properties. Moreover, eventually it becomes possible to correlate such individual chemical characteristics with underlying important chemical processes within a given type of cell and one thereby begins to become acquainted with the cell in dynamic or functional terms.

Examination of the proposal to divide cells into categories based on histochemical characterization reveals at once the importance of selecting significant chemical entities. If, for example, one were to choose the presence merely of protein substances as a common denominator, the group would be too inclusive to be significant. Ideally the distinguishing characteristic should be made sufficiently restrictive so that its presence would reveal a specific cellular activity. As an example one could cite the presence of hemoglobin in erythrocytes. Between the extremes cited there would exist a variety of characteristics of varying degrees of specificity and usefulness.

The usefulness of such a procedure will depend, in the final analysis, upon the number and importance of the new facts which are brought to light and upon the future experiments which this manner of approach suggests. We believe that such an approach relating to the chemical characteristics of diverse groups of cells should illumine many aspects of cellular biology. Although at the present time our technical methods permit only the most fragmentary penetration into this field, we believe that further extension of histochemical methods and correlation will be exceedingly fruitful. We should like, therefore, to present in the remainder of this review a few examples of correlations between histochemical characteristics and cellular activities.

A. *The relation of nucleoproteins to protein synthesis.* Greenstein (1944) has recently reviewed the evidence for nucleoprotein participation in protein synthesis and has discussed the possible mechanisms by which the nucleoprotein exerts its controlling influence. It is consequently unnecessary here to dwell extensively upon the evidence supporting the hypothesis that nucleoproteins are enzymes which regulate protein formation or are templates upon which particular kinds of proteins are fashioned. Instead, we prefer to summarize only briefly some of the information presented by Greenstein and to add a few accessory observations which have become available since his article was published.

As discussed in an earlier section of this review, the cytoplasmic basophilia which is a characteristic feature of many different cells widely located throughout the body is now known to be caused by the presence of ribonucleoproteins. Cells exhibiting cytoplasmic basophilia occur in most parts of the body and may be derived from any of the three germ layers, and since they contain nucleoproteins, it is logical to assume that they are engaged in rapid protein synthesis. Furthermore, in certain instances, differences in the rate of protein formation are accompanied by differences in the amount or distribution of the cytoplasmic basophilia. We should particularly like to emphasize this relationship of nucleoprotein concentration to protein synthesis.

The sero-zymogenic glands associated with the digestive system exhibit

basophilic staining in the basal portion of the cell. Basal chromidial substance is a distinguishing feature of the serous portions of the salivary glands, the pancreatic acini and of certain of the glandular cells within the intestinal tract. In these cells, the zymogen granules ordinarily occupy the apical portion of the cell extending downward so that in the resting gland the most basally located granules overlap the region occupied by the basal chromidial substance. Upon stimulation of the gland to secretory activity the cells are readily depleted of their apically located granules whereas only rarely and after intense stimulation do the more basally located granules disappear (Babkin, 1944). Moreover, during secretory activity, the chromidial substance becomes more abundant and may be stained more intensely (cf. Mirsky, 1943).

It is a well-known fact that the early developmental stages of the blood cells are typically basophilic. In the case of the erythrocyte series, basophilia is inversely proportional to the content of hemoglobin within the cell. Young stages are strongly basophilic and contain no hemoglobin, juvenile stages are characterized by diminishing basophilia and an increasing concentration of hemoglobin, whereas mature erythrocytes stain almost not at all with basic dyes and contain an abundance of hemoglobin. A similar relationship obtains between basophilia and the specific granules of the myelocytes; as the granules increase the basophilia declines to complete extinction in the mature cells. Consequently, in both the erythrocyte and the granulocyte series, basophilia declines inversely with the accumulation of the specific cytoplasmic proteins.

The situation in regard to the lymphocyte is somewhat more complex and until recently was less well understood. The cytoplasm of cells of the lymphocyte series is typically basophilic both in early and mature developmental stages. However, the relationship of this fact to the possibility of protein synthesis has only recently been elucidated. A brilliant series of investigations to which contributions have been made by Dougherty and White (1945), Harris, Grimm, Mertens and Ehrich (1945), Kass (1945), and Ehrich and Harris (1945) has demonstrated that during certain physiological states there is a mobilization of lymphocytes from their stores in the tissues, while at the same time in the blood stream a lymphopenia occurs. This paradoxical situation is explainable on the hypothesis that the lymphocytes undergo liquefaction in the circulation, a corollary to this being that they liberate the proteins composing them into the blood stream. Investigations designed to test this hypothesis have shown that parallel to the disappearance of the circulating lymphocytes there is an increase in the gamma globulin fraction of the plasma proteins, and that washed lymphocytes produce, when lysed, a protein the electrophoretic mobility of which is identical with that of the gamma globulin. Furthermore, Ehrich and Harris (1945) have demonstrated that the lymphocytes contribute to antibody production when activated by antigens. In consequence of these extraordinary findings, it now appears that a major function of lymphocytes is to produce globulins which participate in immune reactions. Consequently, the globulin synthesis of the lymphocyte can now be related to the cytoplasmic nucleoproteins which are responsible for basophilic properties of the cells.

The phenomenon of chromatolysis in nerve cells after section of the axon has long been known. This process involves the dissolution and disappearance of the Nissl bodies during retrograde degeneration of the nerve fiber. Subsequent to the degenerative chromatolysis, regeneration of the axon occurs and there is a parallel reconstitution of the Nissl substance. For several reasons, including their basophilic property, Held as long ago as 1895 suggested that the Nissl bodies were nucleoprotein in nature. However, not until recently was the fact firmly established. Gersh and Bodian (1943) showed that the ultraviolet absorption spectrum of Nissl substance was identical to that of nucleoproteins and, further, that both the ultraviolet absorption and the characteristic tinctorial properties were destroyed by digestion in a solution of crystalline ribonuclease. Thus, in nerve cells there is also a correspondence between nucleoprotein and protein synthesis. In the intact nerve cell, which is presumably inactive with respect to protein synthesis, the nucleoprotein is apparently aggregated into clumps. During the process of regeneration, it may be presumed that synthesis of new protein to reconstitute the axon must occur. In this phase there is a dispersal of the cellular nucleoprotein to be followed again by its segregation into clumps at the close of the regenerative process.

Basophilia may also be associated with protein synthesis in certain of the glands of internal secretion. While investigating the human placenta, Dempsey and Wislocki (1944) were struck by the observation that both the syncytial and cellular trophoblast were strongly basophilic in early pregnancy, whereas in later stages the amount of basophilic substance declined until at term only occasional cytotrophoblastic elements retained their basophilic property. By recourse to digestion with crystalline ribonuclease, these authors were able to identify the basophilic substance as nucleoprotein. The correspondence in the amounts of basophilia exhibited in the trophoblast and the amounts of chorionic gonadotropin excreted in the urine during the various stages of pregnancy prompted the suggestion that the basophilic nucleoprotein might be concerned in the synthesis of the hormone. Moreover, because the pituitary hormones are chemically similar to those from the placenta, Dempsey and Wislocki (1944) sought and found ribonucleoprotein to be a prominent cytoplasmic feature of certain cells of the anterior pituitary gland.

Another example of the correspondence between cytoplasmic basophilia and protein formation stems from a comparative study of the placental barrier (Wislocki and Dempsey, 1945). The glandular and surface epithelium of the uterus, just prior to and during gestation, is rich in basophilic substance, greatest in the uterine glands of the pregnant sow, intermediate in carnivores, and least in rodents and in pregnant women. In the sow, the epitheliochorial placenta consists of a simple apposition of the chorion against the intact endometrium. Consequently, all of the nutritive substances obtained by the fetus must either traverse or be formed and secreted by the uterine epithelium. The abundance of the uterine glands in the sow and the great quantities of ribonucleoprotein which they contain suggest that the proteins available to the fetus reach it through the uterine milk which is secreted by these glands. In the

human, on the contrary, the uterine glands are probably relatively unimportant in the nutrition of the embryo, since in this species the uterine lumen is separated from the fetal tissues by the decidua capsularis. Nevertheless, at the time of implantation and for a period thereafter, the paraplacental and subplacental glands are rich in glycogen, lipoids and basophilic substance, suggesting a nutritive rôle of the glands toward the free blastocyst, and the basal peripheral trophoblast (trophoblastic shell and cell columns) for some time following implantation. Rodents, as exemplified by the rat and mouse, occupy a similar position with respect to the availability of the uterine secretions for the nutrition of the fetal adnexa. In these two species the inverted yolk sac, after the dissolution of the capsularis and the parietal wall of the yolk sac, is exposed to the contents of the uterine lumen. The uterine glands of the rat and mouse, beginning with progestation and extending throughout gestation, have an appreciable complement of basophilia. In the cat, however, the uterine glands are intermediate between the sow, on the one hand, and man and rodents on the other, being larger and more active than those of the latter and possessing much more basophilia. In this series of animals the amount of basophilic substance is greatest in the uterine glands of the pregnant sow, intermediate in those of the cat, and least in rodents and pregnant women.

The last example which we wish to cite concerns the presence of basophilic inclusions in hepatic cells. These inclusions are easily demonstrable after staining with pyronin or with methylene blue. Their protein nature was readily established (Pfuhl, 1932) and they were observed to increase in size and abundance after eating. Despite their relation to alimentation, it was demonstrated that they could not represent albuminoid substances because of their solubility characteristics. Recently, Davidson and Waymouth (1944) have demonstrated that these basophilic inclusions in hepatic cells are abolished after digestion in a solution of crystalline ribonuclease. Their nucleoprotein nature and their variation during alimentation have been confirmed in this laboratory (Deane, unpublished data).

The examples cited above come from diverse morphological locations and concern widely different physiological states. Yet they have in common the fact that cytoplasmic ribonucleoproteins are contained within cells which are actively engaged in the synthesis of some kind of protein substance. It is tempting to speculate that since various kinds of proteins are elaborated by the different cells, the corresponding nucleoproteins must also be different either in chemical composition or in spatial configuration. It is to be hoped that in the future more refined methods will permit us to put to critical test the assumption that special proteins must have special templates upon which they are fashioned.

B. *The relation between nucleoproteins and phosphatase.* We have already had occasion in an earlier section to allude to the chemical data which implicate phosphatase in nucleoprotein and nucleotide metabolism. However, the rôle of phosphatase is far from completely understood. Since nucleotides are in themselves organic phosphates, it is reasonable to suspect that their degrada-

tion into nucleosides might be a function of phosphatase, and indeed it may easily be shown in test-tube experiments that nucleotides and nucleic acids may be enzymatically dephosphorylated by phosphatase. On the other hand, an equally reasonable assumption might be that phosphatase serves the purpose of providing phosphate ions from some other phosphorylated substrate for the synthesis of nucleotides. However, whatever the nature of the relationship between nucleoproteins and phosphatase may be, enough histochemical observations are now available to indicate that these two compounds frequently occur together and to suggest that they may be coupled into some kind of a metabolic process.

The dissolution of the Nissl substance during retrograde degeneration and its subsequent reconstitution after regeneration of the nerve fiber have already been mentioned. Bodian and Mellors (1945) have described parallel changes in the concentration within the nerve cells of phosphatases which operate over a wide range of pH. During chromatolysis marked increases in phosphatase occur, the amount accumulating in the center of the perikaryon being sufficient to account for the eccentric displacement of the nucleus which has long served as a reliable indicator of retrograde degeneration. Later, following regeneration of the axon, the accumulated phosphatase disappears as the Nissl substance is reconstituted and the nucleus returns to its original central position.

A sequence of events rather similar to that discussed above takes place in the human placenta during the course of gestation. Dempsey and Wislocki (1945) have described the gradual disappearance of basophilic nucleoproteins from the syncytial trophoblast of the chorionic villi in the course of pregnancy and the gradual appearance of alkaline phosphatase. As the syncytium becomes less basophilic and more acidophilic, parallel preparations reveal increased amounts of alkaline phosphatase.

The two examples presented above demonstrate an inverse relationship between ribonucleoprotein and phosphatase. Such examples are compatible with the hypothesis that phosphatase acts to degrade or destroy the nucleoprotein. However, we shall illustrate in the next paragraph that this inverse relationship does not always obtain, and therefore that the presence of phosphatase does not necessarily preclude the simultaneous presence of ribonucleoproteins.

The cytoplasm of osteoblasts stains intensely with basic dyes, as is readily demonstrable in undecalcified preparations of bone. After decalcification the basophilic property of the osteoblasts is largely lost because nucleoproteins are readily hydrolyzed in solutions of strong mineral acids. Thus the basophilic quality of osteoblasts is best appreciated in preparations of embryonic bones where decalcification before sectioning is unnecessary or at most need be minimal. Quite unlike the nerve cell and placental syncytium described above, the osteoblast contains abundant alkaline phosphatase in addition to its striking complement of cytoplasmic nucleoprotein. The presence of phosphatase in these cells is a fact familiar to many by both chemical and histochemical methods.

The observations cited above illustrate that phosphatases can exist in quite

different relationships with nucleoproteins in different cellular locations. The thought obtrudes itself that osteoblasts may represent a situation in which phosphatase is secreted by cells to become active in the intervening intercellular matrix of nucleoprotein. If this interpretation be valid, the concentration would find explanation in the rapid production and secretion of the protein phosphatase. Definitive conclusions on these points in question cannot be drawn, but it is to be hoped that future investigations will correlate the occurrence of these two compounds in other locations.

C. *The relation between glycogen and phosphatase.* As a result of the intensive studies of carbohydrate metabolism which have been carried out during the past decade, the biochemical events leading to the formation of glycogen are now fairly well understood. It is generally agreed that a phosphorylated intermediary, such as the Cori ester, is a necessary antecedent to glycogenesis. Beginning with this intermediary, the formation of glycogen then proceeds by dephosphorylation of the hexose phosphate which condenses in the specific environment of the animal cell into glycogen. The transformation of hexose phosphate into glycogen requires the activity of a phosphate-splitting enzyme, and of the various enzymes capable of such action, alkaline phosphatase has been most frequently invoked because of its ubiquitous occurrence and because of its demonstrated capacity to utilize the particular substrates which serve as intermediaries in carbohydrate metabolism.

Such considerations as those in the preceding paragraph lead to the thought that phosphatase activity represents the penultimate stage in glycogenesis. It follows that the enzyme should be demonstrable in a region supervening between the source of the sugar and the location in which glycogen is being deposited. This hypothesis lends itself to experimental proof by correlated observations on the presence of phosphatase and the deposition of glycogen.

A series of comparative observations on the occurrence of phosphatase and the deposition of glycogen has been presented by Wislocki and Dempsey (1945). In both the maternal and fetal portions of the placenta in several species of animals it is found that the distribution of these two substances is in good accord with the hypothesis as stated above. Nevertheless, several interesting differences in the spatial organization of the compounds has been noted. In the sow, glycogen occurs only in the fetal chorionic cells but is entirely lacking in the maternal endometrium. In this species, alkaline phosphatase is confined to the maternal endothelium and to the apical borders of the chorion. Thus, the transmission of sugar from the maternal to the fetal tissues involves its passage across two regions in which phosphatase is demonstrable. Since glycogen deposition occurs immediately after its passage across the second region, the thought lies close at hand that here is the region of final dephosphorylation which precedes the polymerization of the glycogen molecule.

Other variants of the anatomical distribution of phosphatase and glycogen occur in still other species, but these various manifestations also bear out the concept enunciated above. In the human endometrium during early pregnancy, phosphatase occurs only in the capillary endothelium, whereas glycogen

is densely concentrated in the epithelium. In the pregnant cat, on the other hand, the deeply situated uterine glands contain both glycogen and phosphatase, and in this case no other deposits of the enzyme occur in the vicinity. Consequently it appears that, although phosphatase generally occurs between the capillary bed and the region in which glycogen is deposited, the anatomical region which contains the enzyme is by no means constant. It is to be hoped that similar correlated observations in other regions of the body will become available.

SUMMARY

The present review describes a selected series of staining reactions which have been applied to tissue sections and which have provided information capable of chemical interpretation. The point of view is expressed that useful chemical characterization of cells and various interstitial substances can often be accomplished by methods which are not highly specific. Furthermore, by the comparison of several different lines of evidence, conclusions may frequently be drawn when a single approach would be inconclusive.

Several reactions which are reasonably specific for particular compounds or groups of compounds are described. Among these are reactions for lipoidal substances such as lipines, steroids and vitamin A, and several enzymatic reactions including those for oxidases and peroxidase, the dopa reaction and phosphatases. The relationships between the substances revealed by these reactions and various physiological states are briefly discussed.

It is pointed out that a system of histological classification based upon histochemical observations will eventually prove useful. Such a system, employing chemical criteria, will be more directly applicable to the analysis of functional phenomena than the presently used morphological system. As examples of this type of approach to histological problems, the relationships of nucleoproteins to protein synthesis, of nucleoproteins to phosphatase activity, and of phosphatase to glycogenesis are discussed.

REFERENCES

ASPLUND, J., U. BORELL AND H. HOLMGREN. Ztschr. f. mikr.-anat. Forsch. **48**: 478, 1940.
ASTWOOD, E. B., J. SULLIVAN, A. BISSEL AND R. TYSLOWITZ. Endocrinol. **32**: 210, 1943.
BABKIN, B. D. Secretory mechanism of the digestive glands. Hoeber, New York, 1944.
BAKER, J. R. Quart. J. Micr. Sci. **85**: 1, 1944.
BAUER, H. Ztschr. f. Zellforsch. **15**: 225, 1932.
Ztschr. f. mikr.-anat. Forsch. **33**: 143, 1933.
BENNETT, H. S. Am. J. Anat. **67**: 151, 1940.
BICHAT, M. Cited from General anatomy, applied to physiology and medicine. 1801. Transl. by George Hayward. Richardson and Lord, Boston, 1822.
BIERRY, H. AND B. GOUZON. C. R. Acad. Sci. **202**: 686, 1936.
BLOCH, B. Arch. f. Dermat. **124**: 129, 1917.
BODIAN, D. AND R. C. MELLORS. J. Exper. Med. **81**: 469, 1945.
BOURNE, G. Quart. J. Exper. Physiol. **32**: 1, 1943.
BRACHET, J. C. R. Soc. Biol. **133**: 88, 1940.

CASPERSSON, T. Arch. f. exper. Zellforsch. **22**: 655, 1939.
J. Roy. Micr. Soc. **60**: 8, 1940.
CASPERSSON, T. AND J. SCHULTZ. Proc. Nat. Acad. Sci. **26**: 501, 1940.
CORNER, G. W. Science **100**: 270, 1944.
COWDRY, E. V. Microscopic technique in biology and medicine. Williams & Wilkins, Baltimore, 1943.
DAVIDSON, J. N. AND C. WAYMOUTH. Proc. Roy. Soc. Edinburgh **62**: 96, 1944.
Biochem J. **38**: 39, 1944.
DEMPSEY, E. W. Endocrinol. **34**: 27, 1944.
DEMPSEY, E. W. AND D. L. BASSETT. Endocrinol. **33**: 384, 1943.
DEMPSEY, E. W. AND G. B. WISLOCKI. Endocrinol. **35**: 409, 1944.
Am. J. Anat. **76**: 277, 1945.
DEROBERTIS, E. AND J. M. GONCALVES. Endocrinol. **36**: 245, 1945.
DESCLIN, L. C. R. Soc. Biol. **133**: 457, 1940.
DIETRICH, A. Verhandl. d. deutsch. path. Ges. **14**: 263, 1910.
DOUGHERTY, T. F. AND A. WHITE. Am. J. Anat. **77**: 81, 1945.
DUBOS, R. J. Science **85**: 549, 1937.
The bacterial cell. Harvard Univ. Press, Cambridge, 1945.
EHRICH, W. E. AND T. N. HARRIS. Science **101**: 28, 1945.
EHRLICH, P. Das Sauerstoff-Bedürfniss des Organismus. Hirschwald, Berlin, 1885.
Die Med. Woche. 15, 1901.
EMMEL, V. M. Anat. Rec. **91**: 39, 1945.
EVERETT, J. W. Anat. Rec. **91**: 10, 1945.
FEULGEN, R. AND H. ROSSENBECK. Hoppe-Seyler's Ztschr. f. physiol. Chemie **135**: 203, 1924.
FEULGEN, R. AND K. VOIT. Pflüger's Arch. **206**: 389, 1924.
FIESER, L. F. Chemistry of natural products related to phenanthrene. 2nd ed. Reinhold, New York, 1937.
FISCHLER, F. Zentrabl. f. allg. Path. u. path. Anat. **15**: 413, 1904.
FLEXNER, L. B. J. Biol. Chem. **131**: 703, 1939.
FLEXNER, L. B. AND R. D. STIEHLER. J. Biol. Chem. **126**: 619, 1938.
GERSH, I. Physiol. Reviews **21**: 242, 1941.
GERSH, I. AND D. BODIAN. Biol. Symposia **10**: 163, 1943.
GOMORI, G. Proc. Soc. Exper. Biol. and Med. **42**: 23, 1939.
J. Cell. and Comp. Physiol. **17**: 71, 1941a.
Arch. Path. **32**: 189, 1941b.
Proc. Soc. Exper. Biol. and Med. **51**: 133, 1942.
GOULD, B. S. J. Biol. Chem. **156**: 365, 1944.
GREENSTEIN, J. P. Advances in Protein Chemistry **1**: 210, 1944.
HARRIS, H. A. Nature **130**: 996, 1932.
HARRIS, T. N., E. GRIMM, E. MERTENS AND W. E. EHRICH. J. Exper. Med. **81**: 73, 1945.
HELD, H. Arch. f. Anat. 396, 1895.
HEMPELMANN, L. H., JR. Anat. Rec. **78**: 197, 1940.
HOLLANDE, A. Bull. Hist. appliq. à la physiol. **1**: 422, 1924.
HOLMES, O. W. Border lines of knowledge in some provinces of medical science. An introductory lecture. Boston, Ticknor and Fields, 1862.
HOLMGREN, H. Ztschr. f. mikr.-anat. Forsch. **47**: 489, 1940.
HOLMGREN, H. AND O. WILANDER. Ztschr. f. mikr.-anat. Forsch. **42**: 242, 1940.
HOROWITZ, N. H. J. Dent. Res. **21**: 519, 1942.
JORPES, E., H. HOLMGREN AND O. WILANDER. Ztschr. f. mikr.-anat. Forsch. **42**: 279, 1937.
KABAT, E. A. AND J. FURTH. Am. J. Path. **17**: 303, 1941.
KASS, E. H. Science **101**: 337, 1945.
KEILIN, D. Ergebn. Enzymforsch. **2**: 239, 1933.

KIRKMAN, H. AND A. E. SEVERINGHAUS. Anat. Rec. **70:** 413, 1938.
Anat. Rec. **70:** 557, 1938.
Anat. Rec. **71:** 79, 1938.
KUNITZ, M. J. Gen. Physiol. **24:** 15, 1940.
LIPPMAN, F. Ann. Rev. Biochem. **12:** 1, 1943.
LISON, L. Bull. d'histol. appliq. à la Physiol. **9:** 177, 1932.
Arch. Biol. **46:** 599, 1935.
Histochemie animale. Gauthier-Villars, Paris, 1936.
MCCALLA, T. M. J. Bact. **40:** 23, 1940.
Stain Technol. **16:** 27, 1941a.
J. Bact. **41:** 775, 1941b.
MCCALLA, T. M. AND F. E. CLARK. Stain Technol. **16:** 95, 1941.
MCCLUNG, C. E. Microscopical technique. 2nd ed. Hoeber, New York, 1937.
MACKENZIE, C. G. AND J. B. MACKENZIE. Endocrinol. **32:** 188, 1943.
MANN, G. Physiological histology. Oxford Press, Oxford, 1902.
MICHAELIS, L. Biol. Bull. **87:** 155, 1944.
MIRSKY, A. E. Advances in Enzymology **3:** 1, 1943.
MITCHELL, A. J. AND G. B. WISLOCKI. Anat. Rec. **90:** 261, 1944.
PFUHL, W. Handbuch d. mikr. Anat. d. Menschen **5:** 235, 1932.
POLLOCK, W. F. Anat. Rec. **84:** 23, 1942.
POPJÁK, G. J. Path. and Bact. **56:** 485, 1944.
POPPER, H. Physiol. Reviews **24:** 205, 1944.
RALLI, E. P. AND I. GRAEF. Endocrinol. **32:** 1, 1943.
REICHSTEIN, T. AND C. W. SHOPPEE. Vitamins and Hormones **1:** 345, 1943.
ROBISON, R. 1922. Cited from The significance of phosphoric esters in metabolism. New York Univ. Press, 1932.
SALTER, W. T. The endocrine function of iodine. Harvard Univ. Press, Cambridge, 1940.
SARASON, E. L. Arch. Path. **35:** 373, 1943.
SCHACHNER, H., A. L. FRANKLIN AND I. L. CHAIKOFF. J. Biol. Chem. **151:** 191, 1943.
SOBEL, A. E. AND H. WERBIN. Fed. Proc. **4:** 104, 1945.
STOWELL, R. E. Stain Technol. **20:** 45, 1945.
SUMNER, J. B. AND G. F. SOMERS. Chemistry and methods of enzymes. Academic Press, New York, 1943.
TAKAMATSU, H. Trans. Jap. Path. Soc. **29:** 492, 1939.
TOLSTOOUHOV, A. V. Stain Technol. **3:** 49, 1928.
UNNA, K., C. V. RICHARDS AND W. L. SAMPSON. J. Nutrition **22:** 553, 1941.
VERNE, J. Ann. de Physiol. **5:** 245, 1929.
WEED, L. H. Carnegie Contrib. to Embryol. **5:** 1, 1917.
WEINER, P. Ztschr. f. mikr.-anat. Forsch. **13:** 197, 1928.
WESTERFELD, W. W. AND C. LOWE. J. Biol. Chem. **145:** 463, 1942.
WISLOCKI, G. B. AND H. S. BENNETT. Am. J. Anat. **73:** 335, 1943.
WISLOCKI, G. B. AND E. W. DEMPSEY. Am. J. Anat. **77:** 365, 1945.
WOLF, A., E. A. KABAT AND W. NEWMAN. Am. J. Path. **19:** 423, 1943.

THE CORONARY CIRCULATION

DONALD E. GREGG[1, 2, 3]

Department of Medicine, School of Medicine, Western Reserve University, Cleveland, Ohio

As in any other vascular bed, the influences controlling coronary blood flow are comprised of both physical and chemical factors. The flow varies directly with the size of the vascular bed, the pressure head and inversely with the back pressure at the end of the system and the viscous resistance encountered in transit through the entire bed. Aortic blood pressure determines the head applied at the coronary orifices; right auricular pressure (and to a smaller extent pressure in the right ventricle) determines venous back pressure. Viscous resistance to flow is determined by the viscosity of the blood and the bore of the blood vessels. The latter may be altered by two factors: passive compression and relaxation resulting from rhythmical myocardial contraction (extravascular support) and active vasomotor changes in the vessels induced by nervous and/or humoral influences upon their intrinsic muscles.

In past studies of the determinants of coronary blood flow, the venous drainage systems of the two ventricles were considered to operate in fundamentally different ways, and local mechanical forces were believed to be more potent in regulating flow than the chemical or humoral influences. There has been a tendency to stress results obtained in situations where, although certain flow determinants were easily and completely controlled, it was neither possible to gauge the physiological normalcy of the preparation nor to determine the influence of the uncontrollable variables. Recent studies in the dog have led to a re-examination of certain beliefs inasmuch as several new anatomical and functional concepts have arisen concerning the coronary venous system, collaterals, the coronary flow response to drugs, to increased cardiac load, and to sympathetic nerve stimulation. Such modification of our working knowledge of the coronary circulation has followed the introduction of new methods and instruments and new and revised experimental procedures. The object of this communication is to consider the development of these and other concepts, and to discuss the respective merits and shortcomings of the various devices and preparations used to study the coronary circulation and the factors which alter coronary blood flow.

Brief of the Anatomical Structures. The right auricle and ventricle receive most of their blood supply from the right coronary artery, and the left auricle and ventricle are supplied largely from the left coronary artery. However, adjacent portions of the two ventricles and the intraventricular septum are

[1] Now at the School of Medicine and Dentistry, The University of Rochester, Rochester, New York.

[2] The author wishes to express his indebtedness and appreciation to Dr. R. E. Shipley, a former collaborator now at the Lilly Laboratory for Clinical Research, City Hospital, Indianapolis, Indiana, who has taken time to review the manuscript and to write that portion dealing with cardiac nerves.

[3] This article was made possible by a grant from the Commonwealth Fund.

supplied in varying amounts by both coronary arteries. The sinoauricular and auriculo-ventricular nodes are usually supplied by the right coronary artery, while the right and left bundle branches are supplied by both coronary arteries; the arborization systems receive their blood from the arteries supplying the myocardium in which they lie (1, 2).

Comparatively few studies have been made of the capillaries of the heart muscle because of the great difficulty of their injection. Wearn (3), using an injection technique for the isolated beating heart, succeeded in injecting the entire capillary bed and observed that the capillaries are parallel to the muscle fibers and that each muscle fiber lies in direct contact with one or more capillaries. During normal growth a relatively constant capillary-muscle fiber ratio is maintained from birth to maturity. During cardiac hypertrophy the muscle fibers enlarge, but the capillaries do not multiply and hence the capillary supply per unit volume of heart muscle is reduced (4, 5).

The venous drainage systems of the myocardium are 1, the superficial or subepicardial which empties into the right auricle and consists of the anterior cardiac veins and the coronary sinus together with associated veins emptying near its mouth, and 2, the deeper system of veins which communicates directly with the chambers of the heart.

The deeper system of veins (2, 6) is made up of the arterio-luminal vessels (arterial communications between a coronary artery and the heart chambers), the arterio-sinusoidal vessels (thin-walled endothelial tubes of capillary structure connecting arterioles with the heart cavities) and Thebesian vessels (channels between the coronary veins or the distal ends of capillaries and the heart cavities). These channels are much more numerous on the right side of the heart.

The superficial drainage system of coronary veins extends over the surfaces of both hearts. The greater portion of the subepicardial surface of the left ventricle is traversed by many small venous branches which merge to form the larger coronary veins. These converge to form the coronary sinus, which, with associated veins emptying near its mouth, constitute the major superficial drainage system of the left heart. In the right heart (and particularly evident in the dog) the small venous branches merge to form the anterior cardiac veins, anatomically the lesser of the superficial venous systems (7). They constitute several good sized venous trunks which lie buried in the subepicardial fat occupying the sulcus between the right auricle and right ventricle and each empties separately and directly into the right auricle a few millimeters superior to the ventricular border of the tricuspid valve. Other smaller veins are invariably present. The anterior cardiac veins, although described by Grant (8), have been virtually ignored by physiologists.

In addition, there are numerous anastomotic channels connecting the coronary veins of auricles and ventricles with each other, with Thebesian channels, and with extracardial veins draining into the superior and inferior venae cavae (9).

In the anatomical sense, the coronary arteries are not end arteries. Small arterial communications normally exist between the right and left coronary

arteries, between branches of the same coronary artery, and between coronary and extracardial arteries. The intra- and extracardial communications, while small, are numerous and collectively have a considerable diameter. With appropriate stimuli both arterial and venous anastomoses may become greatly enlarged (9, 10, 11).

EXPERIMENTAL APPROACHES TO THE STUDY OF THE CORONARY CIRCULATION. The development of adequate preparations and methods and the overcoming of instrumental obstacles is an essential preliminary to the solution of the problems dealing with the circulation. Such consideration should reveal in part the proper basis for the interpretation of past and future studies of the cardiovascular system.

The direct quantitative determination of total coronary flow requires the measurement of flow in all of the coronary arteries or veins. Flow measurements made in only one artery or vein may or may not indicate correctly the directional changes in total coronary blood flow. In general, when experiments are designed to permit measurement of total inflow or outflow, the preparation usually becomes sufficiently removed from the normal as to limit the interpretation of results.

Blood Flow Measurements with Devices Requiring Cutting the Coronary Vessel. Each device must be inserted between the severed ends of a coronary vessel. The instruments may be conveniently divided into those which measure the mean rate of coronary flow and those which measure continuously the phasic changes in the rate of inflow or outflow. From the phasic flow curves mean rate of flow can also be determined.

A. Mean rate of flow. Among the simpler methods for determining mean rate of coronary flow is the use of a mechanical stromuhr or the perfusion of one or more coronary arteries at a constant pressure and temperature from a calibrated perfusion reservoir (12). In the former the perfusing pressure and blood are supplied from the aorta and in the latter from a chamber outside of the animal. The venturimeter (13), orifice meter (14), and Pitot tube (15) are basically the same in operation and record the pressure drop between two points in the flowing stream. The pressure difference, which is proportional to the flow, may be read from water or mercury manometers or may be optically recorded by a differential manometer. The rotameter (16) has been extensively used to record mean rate of blood flow in both the coronary and peripheral circulation. Basically, this consists of a tapered vertical tube within which the height of a metal "float" varies with the rate of blood flow. For continuous optical recording of flow, a small rotameter (17, 18) has been developed which operates in conjunction with an electrical unit. There is also one type of thermostromuhr (21) that permits optical registration of mean rate of flow in cannulated vessels. Its operation is similar to that of other thermostromuhrs and will be considered with them later. Mean rate of flow may also be measured by timing the passage of an air bubble through a glass tube of known length and volume (19, 20). Of these devices the rotameter, Pitot tube, and orifice meter permit continuous optical recording of the flow changes.

B. Phasic changes in instantaneous rate of flow. For the measurement of phasic changes in blood flow the recording device is designed to have a relatively high natural frequency which permits optical recording of the rapid fluctuations in velocity of flow (flow pattern). This instrument records the instantaneous flow at the point of its insertion into the blood vessel. The different kinds of instruments have been considered in a previous review (22) and hence will be only briefly presented here.

Certain of the phasic flow meters require the coronary perfusing pressure to be artificially maintained at a constant level, the coronary vessel being perfused from a reservoir partially filled with blood under the desired pressure. The movements of blood from reservoir to coronary artery (or in the case of a coronary vein, from vein to reservoir) may be indicated by recording the cooling of a heated platinum wire placed in the air in the mouth of the reservoir (23). This device is unable to record rapid changes in flow and is also unable to indicate periods of retrograde flow (back flow). The perfusion reservoir may be closed, blood put under the desired constant perfusing pressure, and a sensitive calibrated optical capsule used to record the slight drop in chamber pressure (24). The calibrated slope of the declining pressure curve gives the rate of blood flow and the total drop over a period of time represents total flow. This instrument has an adequate frequency and records with no appreciable lag. However, because of the maintenance of a constant perfusing pressure, the use of either method permits deductions only with reference to the factors regulating the resistance to coronary flow.

Phasic changes in blood flow may also be determined by methods in which the coronary artery is perfused by the normally pulsating aortic pressure. Optical recordings have been made of the movements of a globule of mercury (25) or toluene (26) interposed in the coronary arterial stream. There is considerable lag inherent in both methods. The strömborste (27), as yet used only in the carotid artery and the aorta, has a high frequency and should have a place in the studies of the coronary circulation. This device consists of a quartz fiber mounted between two electrodes and enclosed in a glass cannula inserted between the cut ends of the blood vessel. Changes in stream velocity cause movements of the fiber which vary the resistance between the electrodes, thereby causing a current change through the recording galvanometer. Of the phasic flow meters, the orifice meter (14, 28) has perhaps been most highly developed and extensively used to measure phasic flow in peripheral vessels as well as in the coronary circulation. It was applied to the study of coronary flow with the hope that by an analysis of part or all of the flow pattern, the effect upon blood flow caused by cyclic changes in intravascular volume associated with volume-elastic changes and extravascular compression might be separated from the effect of intrinsic vasomotor changes in the coronary vessels. As will be indicated later, this has been only partially realized.

All of the aforementioned devices have certain advantages and disadvantages; some have superior technical aspects. For example, the rotameter, orifice meter, and strömborste permit continuous recording of flow changes. All of

the instruments limit flow to a variable extent and have the disadvantage of requiring their insertion into the severed vessel of an operated and anesthetized dog to which anticoagulants have been given. Such procedures must induce a variable and unknown degree of insult to the cardiovascular system and its associated nervous and metabolic mechanisms; hence, the interpretation of the experimental results in terms of occurrences within the normal animal is necessarily restricted.

Measurement of Blood Flow with Intact Coronary Vessel. These methods do not require severing of the coronary vessel and in most instances no anticoagulant is required. In some cases flow measurements may be made in the unanesthetized state.

The electromagnetic flow meter (29, 30) is used to measure mean or phasic rate of flow. As yet it has not been used to measure coronary blood flow. Compared to the orifice meter flow curves, the patterns recorded with this instrument are somewhat "damped." This device utilizes the principle that an EMF is induced in the blood stream as it flows through a magnetic field. The method would appear to be adequate for the measurement of mean rate of flow.

Flow through the coronary vessels as well as other vascular beds has been determined by several types of thermostromuhrs. Most of these instruments record the temperature difference of two thermal junctions applied to an unopened vessel when the vessel and blood between them are heated by means of high frequency current (31) or in direct current (32, 33). Bennett, Sweet and Basset (21) have devised a thermostromuhr onto which the severed blood vessel is tied and through which the passage of blood cools a heated thermocouple mounted on the end of a spicule projecting into the lumen of the device. Schmidt and Hendrix (34) first suspected that the thermostromuhr was subject to a number of possible sources of error. Subsequent extensive tests of this type of instrument (21, 35, 36) made *in vivo* and *in vitro* have indicated that the galvanometric deflection varies not only with the rate of flow but with a variety of factors such as the artery used and its degree of stretch, the degree of angulation of the unit with respect to the artery, the presence of periods of "zero flow" or retrograde flow (back flow) in the flow pattern, the immediate environment as regards composition and movements of intra- and extravascular fluid near the artery, and finally the viscosity of the metered fluid.

The introduction of error by the appearance of, or a change in, the amount of retrograde flow (back flow) in the flow pattern was found to be of particular importance since flow patterns (28, 37) of all arteries thus far recorded with the orifice meter either normally contained back flow or it could be induced by the administration of drugs or by other procedures. The thermostromuhr is particularly sensitive to such changes in that when the actual intra-arterial mean rate of flow decreases and short periods of back flow appear in the flow pattern, the thermostromuhr indicates a large increase in flow (35). Although it would appear to be a very difficult task, it is hoped that ways and means will eventually be found to eliminate the gross errors inherent in this method, since it is the only method of practical design which may be used in unanesthetized animals.

Machella (38) measured phasic coronary flow by recording the cooling of a heated wire threaded through an otherwise intact coronary artery. The flow pattern is somewhat "damped" as compared to that recorded by the orifice meter and no back flow components are present. On a theoretical basis alone the instrument would be incapable of discriminating between forward and back flow and would be particularly inaccurate at low velocities of flow (21).

Changes in phasic coronary flow have been predicted by the so-called differential pressure method (39, 40). More recently with the aid of the rotameter, orifice meter (41) and constant pressure meter (42) researches have demonstrated that such a method may not indicate the directional shifts in flow. In other instances, directional changes may be correctly indicated, but the magnitude of the flow change is not even remotely represented. While such a procedure has added to our knowledge concerning flow determinants it must be discarded as an index of coronary flow.

EXPERIMENTAL PREPARATIONS. Since the time of Starling much information has been gained from experimental preparations in which the coronary circulation has been studied with the heart in various degrees of separation from the rest of the animal. These preparations include the heart-lung (43, 44, 45, 46), the isolated heart with artificial lung (47), and the fibrillating heart (48). In general, the experimental findings have been considered significant and applicable to the normal heart beating *in situ*. The chief advantage in using such preparations is that the several variables such as peripheral resistance, arterial and venous pressures, and cardiac output can be more easily controlled and the coronary flow more easily measured. It is, of course, justifiable to determine what *can* happen with the hope of ultimately learning more about what *does* happen in the intact animal. However, any great hope of success for this approach seems incompatible with the practice of removing the heart from its normal environment, work, and physiological controls and substituting an artificial set of new, and, in many instances, highly abnormal conditions. There are many reasons for believing that much of the great bulk of work with these preparations has only a very limited significance and has yielded results rather far removed from the normal occurrences. The apparent confidence that artificial regulation of the mechanical or tangible variables implies a well controlled (although abnormal) experiment is not warranted. Important unrecognized changes can and do occur in the more intangible mechanisms of physiological function which invalidate many of the interpretations attempted. When compared to studies with hearts beating *in situ* and under more nearly normal physiological conditions, the contour and values of the pressure curves, the coronary A-V O_2 difference, the volume of coronary inflow, the distribution of coronary outflow, and the effects of different variables upon it, such as changes in cardiac output, cardiac load and coronary vein occlusion, are grossly different.

Final judgment as to the validity of many accepted conclusions based on such experiments must await extensive reinvestigation in which utilization is made of preparations with greater functional similarity to the normal heart. In any event, the continued preferential use of such preparations does not seem justifiable when more nearly normal preparations are now available.

Studies have been made with the exposed heart beating *in situ* in the anesthetized dog, in which the coronary vessels were perfused from the aorta, using many of the devices described previously. This is a closer approach to the normal animal, involving less insult to the animal's own nervous, metabolic, and cardiodynamic mechanisms. Some of the studies presumably indicate with reasonable exactness changes which occur in the normal animal. Even so, the observations can be accorded only limited interpretation.

The ultimate objective is the quantitative measurement of coronary flow and the study of its determinants in the normal, unanesthetized animal. Unfortunately, no device has, as yet, been developed which is adequate for this purpose.

METHODS FOR ESTIMATING DETERMINANTS OF BLOOD FLOW CHANGES. The determinants of coronary flow were indicated to be the pressure head, viscous resistance encountered in transit through the bed, and the pressures in the heart cavities at the end of the flow circuits. Changes in aortic or central coronary pressure, venous, auricular, and ventricular pressure can be readily determined by the use of appropriate pressure manometers (49, 50). Viscous resistance to flow is governed for the most part by the mean calibre of the vascular bed. Changes in this, the gross vasomotor state of the bed, can be recognized under certain conditions by correlating, as a simple ratio, the mean flow value with the corresponding mean applied pressure value. However, the application of such an index is limited and should be made with caution. Such an index does not take into account viscosity effects which, in a dynamic system, may give rise to considerable variation in the relationship of mean pressure to mean rate of flow, for which appropriate *in vivo* correction curves are not obtainable (28). Nevertheless, the vasomotor state of a bed may be regarded as having changed when mean rate of flow and mean blood pressure undergo considerable change in opposite directions. Only partial success has resulted from attempts to identify and estimate separately the changes in two important variables which control the calibre of the coronary vessels, namely, extravascular compression (support) and active vasomotor changes in the vessels induced by nervous and chemical influences upon their intrinsic muscles.

The problem of determining the magnitude of extravascular support has been approached in different ways. 1. It has been suggested that intramural pressure may be taken as a measure of extravascular compression. Phasic pressure pulses recorded from within the myocardium of the left ventricle by means of imbedded fluid pockets or arterial segments (51, 52, 53), are greater in the deeper lying myocardium, generally exceed the aortic or ventricular pressure simultaneously recorded, and are believed (51, 52) to approximate the correct value for intramural pressure. However, *in vivo* and *in vitro* tests have shown that the relationship of the recorded intramural pressure to aortic pressure varies greatly with the level of inflation of the segment or pocket and with movement or distortion of the segment, and that if a vessel segment is placed within the left ventricular cavity the pressure generated within it may be 2–4 times the intraventricular pressure. Mechanical protection of the segment reduces but does not eliminate the discrepancies (53). While the use of pro-

tected segments may possibly indicate directional changes, accurate quantitative measurement of intramural pressure is not possible by such methods. 2. The maximum pressure transmitted during systole from the myocardium into a coronary artery, peripheral to the point of temporary occlusion of that artery, has been recorded in many dynamic states and has been used as an index of extravascular compression (14). In such a system elevation of the intracoronary vessel pressure during systole is related not only to the external compressing force of the myocardial fibers, but to the relative fullness of the bed and the ease with which the trapped blood can run off peripherally through the capillaries (vasomotor state). In order that changes in extravascular compression may be accurately indicated it is necessary that the vasomotor state of the bed and its relative fullness remain unchanged. Since it is not possible to control or predict the constancy of these variables the method is to be discarded.

As yet unsolved is the problem of determining the relationship of blood flow to active vasomotor changes, irrespective of whether the effect on intrinsic muscles of the coronary vessels is mediated through the blood stream or is secondary to metabolic changes in the surrounding myocardium. As mentioned earlier, one may be certain of a change in vasomotor state whenever mean rate of flow and mean central coronary pressure change considerably in opposite directions. However, it has been claimed that by analysis of phasic inflow curves changes in vasomotor state and extravascular support can be determined. Critical points on flow curves have been selected in late diastole and systole at which the rate of change of volume elastic and myocardial compression forces are presumed to be minimal (14). At the diastolic point extravascular forces are at a minimum and the rate of flow reflects the vasomotor state of the coronary bed; at the systolic point extravascular support is maximal and the flow is said to reflect the combined effect of myocardial compression and the existing vasomotor state (54, 55). At these points the ratio of rate of flow to the simultaneously existing aortic pressure is determined. A shift in the diastolic ratio is taken to represent active constriction or dilatation of the coronary bed. Changes in the extravascular compressing force are determined by comparing the systolic and diastolic ratios. Such deductions require the necessary proof of the ability of an indirect analysis to indicate correctly the actual occurrences. Some of the assumptions upon which the validity of the method must necessarily depend are that 1, a period exists in which, throughout most of the myocardium, contraction of the myocardial fibers on the lumina of the whole coronary bed is exerted essentially simultaneously; 2, rates of flow taken at the critical points represent flow through the myocardium; 3, rate of flow is linear in both systole and diastole with respect to pressure. Until these are verified experimentally such a system of interpretation cannot be regarded as having a sound hydrodynamic basis.

Arterial and Venous Collaterals. In the study of collaterals the coronary arteries may be occluded abruptly or gradually. For chronic experiments an artery may be tied off abruptly, but more gradual constriction of the lumen may reduce the incidence of ventricular fibrillation, minimize infarction, and aug-

ment collateral development. The use of adjustable mechanical or osmotic clamps or the application of tight-fitting bakelite rings or cellophane bands to the artery will ultimately lead to complete coronary occlusion (56, 57). Unfortunately, by none of these methods can the time of complete occlusion be known, nor can the per cent reduction in flow be predicted even if the extent of local reduction in vessel lumen were known. As in other vessels, the effectiveness of a given localized constriction in reducing flow may be large or small and will vary in inverse relation to the peripheral resistance of the vascular bed and lumen area of the constricted segment and in direct relation to the axial length of the constricted area, flow velocity and blood viscosity (58).

No entirely adequate method is available to measure collateral coronary blood flow in acute or chronic experiments after occlusion of a coronary artery. An indication of intercoronary collateral development can be demonstrated experimentally by measurement of retrograde flow from a severed main coronary branch. In acute and chronic occlusion experiments, with otherwise normal hearts, the retrograde flow of arterial blood is at first small, but this increases within a few hours, becomes considerably greater within a few days, and approximates in days to weeks the preocclusion value for the normal rate of inflow to that branch (59). This process is slow compared to that found in other less vital body regions. Also, following chronic occlusion of a coronary artery, the pressure curve recorded in the artery peripheral to the occlusion may approximate that in the central coronary artery, and gross evidence of development of interarterial collaterals can be shown roentgenographically with the injection of a radiopaque medium (60, 61). If intracardial sources of collateral blood supply are insufficient, extracardial anastomoses also become evident and sizable extracardial communications can be demonstrated. Under the most favorable conditions, the major portion of the coronary arterial system can be occluded with minimal or no infarction in those dogs that survive (66).

Many dogs do not survive more gradual experimental coronary occlusion. In addition to gross coronary insufficiency, it has been suggested that the ultimate cause of death, generally ventricular fibrillation, probably stems from small zones of myocardial anoxia which are highly irritable, and is not necessarily related to a failing heart or gross coronary flow inadequacy. Hence, if small amounts of blood via collaterals could be made to enter potential "trigger zones," it might prevent ventricular fibrillation and eliminate scar tissue formation through muscle preservation. Accordingly, attempts have been made to subsidize the natural ability of growing tissue to develop collaterals. Application of various tissue grafts have been used to form extracardial communications; epicardial or pericardial inflammation has been produced to foster intercoronary and extracoronary communications through mechanical or chemical irritation (62), or by the induction of cardiac venous stasis (63, 64, 65). Opinion based upon physiological experiments is divided as to the functional rôle played by such channels (66, 67, 68, 69). In any event there is a considerable reduction in mortality with all of these procedures, but with venous occlusion no definite difference in the size of infarct can be demonstrated (63).

Distribution of Coronary Outflow. Direct measurement of coronary venous

outflow in isolated hearts and heart-lung preparations has shown that the coronary sinus and associated superficial veins emptying near its mouth constitute the major venous drainage system of the left coronary artery. In such preparations the coronary sinus flow is taken to represent roughly from 50–65 per cent of the total coronary inflow in different cardiodynamic states (46, 70). Since simple and accurate methods for measuring coronary inflow were not available, such measurement of sinus flow has been widely used (with the heart beating *in situ*) to indicate quantitative and directional changes in flow. Much controversy has arisen over the validity of this method, but the arguments are largely supported by observations made with the heart-lung or isolated heart preparation (46, 71, 97, 48, 73). The fact that, with the heart beating *in situ*, coronary sinus flow increases with elevation of right ventricular pressure, has in the past been taken to mean that coronary venous outflow is merely redistributed between the coronary sinus and Thebesian vessels in favor of the coronary sinus (15). This is offered as an argument against the use of coronary sinus flow as an index of coronary inflow. It is more probable that the augmented sinus flow is predominantly the result of the increased inflow in both right and left coronary arteries which occurs (but was not measured in these experiments), especially since Thebesian flow is much less than previously supposed. Experiments showing that the inconstancy of the coronary sinus fraction applies also with the heart beating *in situ* have not been reported. It is probable that a more consistent and reliable evaluation of coronary inflow-coronary sinus outflow relationship could be made with a less artificially controlled preparation which has been permitted to retain a more nearly normal state of hemodynamics. However, from a practical standpoint the issue would seem of little import since there are now available adequate and simple methods of directly measuring coronary inflow in a somewhat more reliable preparation.

From the use of similar preparations (heart-lung and isolated heart) it has been claimed that the Thebesian vessels drain most of the right coronary blood into the right ventricular cavity (73), *i.e.*, the functioning of the drainage system of the right heart was presumed to be considerably different from that of the left heart (48, 73). It is significant that in such studies it was assumed that all coronary venous blood not collected in the coronary sinus and appearing in the right ventricle must have come largely if not entirely from the Thebesian vessels. In a recent paper Katz *et al.* (74) reported a reinvestigation of the problem in which the coronary venous drainage into auricles and ventricles was separated by an instrument made to expand in umbrella-like fashion at the level of the A-V valves. A considerable drainage into the right ventricle was again found and regarded as evidence that "a significant right ventricular Thebesian drainage occurs—which is an important element in coronary drainage." It would seem extremely hazardous to attach particular physiological importance to this finding since 1, it was apparently not noted what effect, if any, the presence of the instrument had upon the outflow from the anterior cardiac veins, the mouths of which open just superior to the ring of the tricuspid valve, and 2, the experiments were performed by perfusion of serum-saline mixtures through the coro-

nary vessels of an excised, non-beating heart, which conditions do not retain a reasonable similarity to those of a living, beating heart, still exercising its normal function in the animal.

Recent work has demonstrated that with the heart beating *in situ*, the venous drainage of the right coronary artery is largely by way of the superficial anterior cardiac veins to the right atrium, under both normal and altered dynamic states. The functional importance of these veins had been virtually ignored by previous investigators. Experimentally, the direct cannulation and measurement of flow in only the major anterior cardiac veins has permitted recovery of from 50 to 92 per cent of the blood entering the right coronary artery. (In addition, a certain fraction of blood from the latter drains into the coronary sinus.) These percentages are considered minimal, since for technical reasons no attempts were made to measure the outflow from the smaller anterior cardiac veins, the right auricular veins, and the many small veins of the right ventricle emptying into the great cardiac vein, all of which constitute channels of exit for blood supplied to the right heart by the right coronary artery (7). These findings in the beating heart make completely untenable the conventionally accepted belief that nearly all or even most of the right coronary inflow drains by way of the Thebesians into the right heart. Most of the coronary venous blood from each heart appears to drain by way of its own system of superficial veins. It follows that the studies of cardiac venous channels, flow partitions, and Thebesian drainage, which have been made in the past without measurement of flow in the anterior cardiac veins should be subjected to a critical revaluation before their associated conclusions can be accepted.

Luminal Vessels and Their Function. Histological studies, including serial sections, reconstructions, dye injections and perfusion experiments of the isolated heart (2) have been used to demonstrate the presence and distribution of Thebesian and luminal vessels through which the coronary vessels communicate with the atrial and ventricular chambers of the heart. These channels are much more abundant in the right ventricle. As already indicated, experimental studies in the beating heart *in situ* suggest that these luminal vessels may function in only a limited fashion in the undisturbed normal heart and to an extent very much less than previously maintained. Actually, although many experiments have been devised to determine if, when, in which direction, and to what extent blood flows through these channels, none has been sufficiently conclusive to justify more than a speculation as to whether they ever function in the normal heart beating *in situ*. The critical objections to most of the experiments can be resolved into either: 1, the false assumption that all coronary venous blood entering the right auricle and right ventricle exclusive of that entering from the coronary sinus, has its origin in the Thebesian vessels, thereby ignoring the considerable venous flow from the anterior cardiac veins, and/or 2, the establishment by artificial manipulation of decidedly abnormal pressure gradients between the coronary vessels and chambers of the heart for the purpose of demonstrating what are implied to be the normal physiological magnitude and directional changes in Thebesian flow.

It is of interest to examine the possible rôle of the Thebesian and luminal vessels in hearts in which either the coronary arterial or cardiac venous system is insufficient. It is conceivable that the luminal vessels could be important, particularly if they could serve as arterial channels from the left ventricle to the myocardium during left coronary artery occlusion or as venous channels for the whole myocardium in the presence of extensive superficial vein occlusion. Known observations are conflicting and any conclusion is difficult. Regarding the first situation, with temporary functional separation of one or all coronary arteries from the aorta, no blood flow from the ventricles into the coronary arteries or superficial venous system could be demonstrated and the hearts did not survive (9, 75). Essentially complete occlusion of the coronary arteries in humans has been found at autopsy (76) but the presence and extent of development of extracardial arterial collaterals was not recorded. When dye is injected into the right ventricle in acute experiments, extensive capillary injection occurs if right ventricular pressure is artificially made to exceed left ventricular pressure. However, the anterior cardiac veins were not excluded as a vascular portal of entry for the dye (77). Regarding the second situation with acute closure of all grossly visible anterior cardiac veins or the coronary sinus, no great reduction in right or left coronary artery flow is usually observed. The heart, following the acute closure of both the coronary sinus and anterior cardiac veins, becomes exceedingly hemorrhagic and tends to become progressively weaker, although such hearts may survive up to two hours. Dogs in which both superficial venous systems have been chronically occluded in a two-stage operation have survived for a period of months. This is suggestive evidence that all venous drainage occurred by way of luminal vessels. However, that significant drainage occurred through such a route could not be verified, since at postmortem examination these hearts exhibited numerous superficial cardiac veins of considerable size which were not previously evident, and several large extramyocardial venous anastomoses, the aggregate cross-section of which might well have been adequate for the venous drainage of the entire heart (9). Until such intra- and extracardial arterial and venous collaterals which appear with coronary arterial or venous ligation have been successfully removed or otherwise excluded as flow channels, a statement as to the utilization of Thebesian channels in diseased hearts is not warranted.

Responses of Coronary Circulation to Augmented Load. Of particular interest has been the subject of the coronary flow response to changes in the working load of the right and left hearts. In experiments with the heart-lung and comparable preparations (but excluding those in which the vagi are intact), elevation of right ventricular pressure by constriction of the pulmonary artery or elevation of left ventricular pressure by aortic constriction (coronary perfusing pressure kept constant) has been demonstrated to cause a reduction in blood flow to the myocardium of right and left hearts respectively (78, 79, 80). The flow decrease is attributed to the dominant effect of the direct mechanical inhibitive action of the increased vigor of the heart and/or the establishment of an unfavorable pressure gradient, especially for the right coronary artery drain-

age via the Thebesian channels into the right ventricle. Such a response would seem to be one of poor economy and certainly would not be advantageous to the heart when under an increased load. However, in recent studies with the heart beating *in situ* and under more nearly normal physiological conditions, with vagi and sympathetics intact or severed, directly measured right and left coronary inflow increased appreciably upon elevation of right or left ventricular pressure (81, 82). In both instances cardiac venous outflow in both the coronary sinus and the anterior cardiac veins increased greatly (83). Measurements of cardiac output and O_2 consumption in the presence of an augmented load have demonstrated an increase in work and metabolism of the respective ventricles. Under these conditions and in consideration of the increased coronary flow, the two ventricles have at their disposal a compensatory means by which their blood supply can, at least in part, be adjusted to their work and metabolic requirements. Although the exact mechanism for the flow increase could not be identified with certainty in these experiments, it seemed probable that the coronary dilatation arose from an increased local production of metabolites and/or the creation of local relative anoxia due to the increased oxygen utilization. The sustained flow increase is taken to indicate a dominant influence of coronary dilatation over the mechanical flow inhibiting effect of an increased extravascular support. As already indicated these results are diametrically opposed to those of other investigators using less normal physiological preparations.

While an increase in coronary flow is the net response to increased cardiac work, it can be shown to be antagonized by a flow-limiting mechanism (81). If, in the presence of an adequately maintained central coronary pressure, the pulmonary artery constriction and release is abrupt, the sustained flow increase described above may be preceded and followed by a transient period of flow reduction and elevation respectively, whose beginnings are coincident with the time of sudden constriction and release of the pulmonary artery. The abrupt changes in coronary flow coincident with the onset of elevation and reduction in ventricular pressure are regarded as indications of a temporary separation of the influence of change in extravascular compression upon coronary flow. This conclusion is made with reasonable justification since the increase or decrease in mechanical compression of the coronary vessels should occur simultaneously with the increase or decrease in intraventricular tension, while the slower physiological metabolic and vasomotor responses must necessarily lag somewhat behind their respective exciting causes, *i.e.*, the sudden changes in cardiac work. The initial decrease in flow can be attributed to the dominant influence of augmented extravascular mechanical compression of the coronary vessels. The transient flow increase following abrupt lowering of intraventricular pressure is a rough index of the extent to which flow had previously been retarded by the augmentation of extravascular compression.

If such experiments are intentionally prolonged (4–5 hrs.), the right coronary flow response to pulmonary artery constriction is reversed and ultimately a sustained decrease in flow will be observed followed by a return to the control

level upon release of the pulmonary artery. Thus, in the latter stages of an experiment during which the condition of the heart *in situ* has progressively become far removed from normal, the flow reducing effect of increased extravascular support is dominant over whatever flow promoting mechanism the heart has retained. As measured under these conditions and by others using heart-lung preparations, a decrease in coronary flow cannot be regarded as the "normal" response to an increased cardiac load.

In the foregoing it was indicated that the coronary flow response to an augmented resistance load is the same for right and left hearts. In all types of heart preparations central coronary artery pressure elevation *per se* is an effective means of increasing left coronary inflow. Hence, in those instances in which the aorta was constricted peripheral to the coronary ostia, the blood supply of the left heart was additionally increased by the elevated central coronary pressure. For this reason the left heart will operate at an advantage over the right heart in its ability to function with an increase in resistance load.

An increase in the venous return to the heart beating *in situ* augments the coronary flow only mildly when aortic pressure is unchanged and to a much greater extent when the aortic pressure is permitted to rise (41). Here, as with increased resistance loads, the increased coronary flow presumably results from augmentation of cardiac metabolism and its associated dilator mechanisms. More recently Katz *et al.* (45) have concluded that the coronary flow increases with augmented cardiac output. The explanation advanced is that the coronary vessels are passively dilated and their resistance to flow decreased as a mechanical adjustment to increased cardiac work and energy expenditure. This concept, at present, lacks the necessary explanation of the physical mechanism by which an increased cardiac output without an increased aortic pressure can of itself cause passive mechanical dilatation of the coronary bed.

Flow Patterns. Phasic flow curves have been recorded with the orifice meter in the coronary arteries and most major peripheral arteries under a great variety of dynamic states. It is desirable to know to what extent such curves may represent events in the coronary bed within the myocardial walls. For the peripheral flow patterns a method of analysis and interpretation has been presented, by means of which the probable determinants of, and their interrelated influences upon, phasic rate of inflow were discussed (37, 84). The coronary flow patterns are more complex. In experiments with anesthetized dogs, left coronary flow patterns show flow peaks in systole and early diastole separated by short periods of reduced rates of flow or back flow during the isometric contraction phase, late systole, or early diastole (14, 22, 85). The right coronary flow pattern is not grossly dissimilar (82). The effects on the flow pattern of changes in blood pressure, heart rate, resistance load, of cardiac nerve stimulation, valvular deficiencies, and vasomotor drugs are multiple and have been separately reported (22, 41, 54, 55, 82, 85, 86, 87, 88). Unfortunately, the analytical approach for peripheral patterns does not encompass all the factors which comprise the determinants of coronary flow patterns. In this case a comparable analysis is not possible because venous outflow is markedly pulsatile

and the arterial flow pattern is greatly influenced by compression and relaxation of extrinsic origin (extravascular support). A method of analysis using critical points on the flow curve has already been considered. It is believed that no adequate system of analysis is as yet available for coronary flow patterns.

Drugs. Drugs may be effective in changing the myocardial blood supply through alteration of the blood pressure and/or the vasomotor state of the normal and collateral coronary vascular channels. Certain of the drugs in addition alter blood flow by modifying the vigor of cardiac contraction. Through lack of adequate methods great confusion still exists as to the mechanism of drug action, and in most instances no definite statement can be made.

Pitressin decreases coronary flow by all methods of study and constricts the coronary bed, *i.e.*, flow decreases in the presence of an increased central coronary pressure (55, 87). It is believed that this drug acts by direct effect on the coronary vessels, but a decrease in cardiac metabolism and vigor of contraction has not been ruled out as the cause of the flow decrease.

The flow response to the xanthines and nitrites is similarly consistent by all methods of study (54, 87, 88). The conclusion is inescapable that this group of drugs exerts a dilating action on coronary vessels, *i.e.*, coronary flow increases in the presence of a normal or reduced central coronary pressure. A similar flow response occurs with papaverine when injected either intra-arterially or intravenously in dogs (89). Whether these drugs directly affect the vessels or cause the flow alteration by inducing cardiac metabolic changes is not known.

In most preparations including the heart beating *in situ* epinephrine increases coronary flow (55). There is also an associated increase in vigor of contraction and cardiac metabolism. The flow increase may be regarded as the net effect which results from an augmented extravascular support, tending to reduce flow, a metabolic dilator effect tending to increase flow, and whatever direct dilating effect the epinephrine may have at the time on the coronary vessels. It is difficult to determine the respective magnitude of each separate effect. Regarding its dilating effect *per se* on the coronary vessels, evidence is conflicting. Green *et al.* (55) found that following injection of epinephrine into the left coronary artery, the coronary flow increased simultaneously with elevation of blood pressure and increased vigor of contraction, while Shipley and Kohlstaedt (89) indicate that this is preceded by a transient period of flow increase without change in blood pressure. This initial flow augmentation might then indicate a direct dilator action on the coronary vessels although, even here, an early metabolic influence cannot be ruled out.

Nervous Influences. The control of the coronary circulation by parasympathetic and sympathetic nerves has been a subject of much study. The nerve supply to the heart is abundant although the anatomical separation of the two systems is somewhat difficult, particularly within the heart. Wollard (90) has reported the presence of nerves, the terminal filaments of which pass among the muscle fibers and within individual muscle cells, and Nonidez (72) has shown the presence of parasympathetic fibers ending in the myocardium of the auricles and ventricles. These findings suggest the possibility of nervous control over

myocardial contraction. The abundance of nerves found in proximity to or ending on the coronary vessels implies a probable vasomotor function.

For the most part the nervous influences on the coronary circulation have been studied by observing the coronary flow responses following electrical stimulation or severance of the nerves. Although such procedures are not paralleled by normal occurrences in the animal, the observed responses are presumed to indicate the functions which the nerves are capable of exercising in the intact animal. Further difficulty in interpretation results from the fact that the specific effects upon the heart muscle and on the coronary vascular system are experimentally inseparable and only the net effect may be observed. Differences in methods and preparations are additional variables which may account for the discordant results of different investigators.

The vagus and sympathetic nerves are reported to contain both dilator and constrictor fibers according to different investigations (cf. reviews by Wiggers, and McDowall (91, 92)). In general the so-called inhibitor and constrictor fibers are said to predominate in the vagus while the dilator and accelerator fibers are found in the sympathetic nerves. The evidence for the constrictor effects has usually been the observation that abolition of the parasympathetic pathways (by mechanical or chemical means) results in an augmentation of heart rate or coronary flow (43). The normal presence of a tonic constrictor and inhibitor action is therefore implied.

Studies of action potentials and the effects of faradic stimulation of the sympathetic nerves to the heart indicate that they probably function in maintaining and augmenting the rate and vigor of the heart beat. (Cf. 85, 93, 94, 95.) Although coronary flow increases with sympathetic stimulation, the work and metabolism of the heart are also increased to such a degree that the accompanying coronary dilatation could be attributed to the increased local production of metabolites (95).

Katz and Jochim (96) have used the fibrillating heart in order that changes in blood pressure, vigor of cardiac contraction, and other variables could be excluded and purely vasomotor responses observed. Their results are not conclusive evidence for the existence of a direct sympathetic vasomotor influence inasmuch as an increase in vigor of fibrillary contraction and cardiac metabolism was not excluded as the primary response which could secondarily give rise to vascular dilatation.

It would seem most difficult to establish and identify conclusively by experimental means the separate effects of nervous influences upon the myocardium and coronary vessels because the physiological functions of these structures are so intimately related that their individual responses may be secondarily modified, each by the other.

COMMENT

The cardiovascular organs form a highly specialized system in which the coronary circulation plays a singularly important, although indirect, rôle in maintaining the organism in a normal hemodynamic state. This fact alone

has amply justified the numerous studies of the coronary circulation and should continue to stimulate many future investigations.

In the present paper no attempt has been made to organize a comprehensive review of all work done in the field and many topics of limited academic interest have been omitted. In addition, the scope of this paper does not permit the thorough consideration of individual topics which may be found in separate reviews. On the other hand, more emphasis has been placed upon the significance of studies with respect to their importance in revealing physiological occurrences which may have a parallel in the normal animal or human.

Because of technical difficulty and, more often, because of the impossibility of making direct and positive observations of the dynamic occurrences in the heart, the investigator must often accept the alternative of observing the responses to variables introduced in an experimental preparation which is considerably removed from the normal state. Except for the anatomical studies the reported observations dealing with coronary flow, its minute volume, distribution and ultimate drainage, its response to drugs, nervous and humoral influences, to changes in aortic pressure, peripheral resistance, cardiac output, cardiac work and metabolism, all are significant only in that they necessarily apply to the preparation and conditions under which the observations were made. All too often the results obtained with a decidedly abnormal preparation ultimately come to be regarded as events which can and do occur in the normal animal or man.

Various concepts concerning the coronary circulation have frequently been revised as new and improved instruments and preparations have been developed. Much of what has been reported in the past should be discarded as more accurate methods are used for making the same physiological studies. Observations made in the open-chest, but otherwise intact dog, must certainly be as cautiously interpreted as those made with any other preparation. However, with the use of accurate methods and instruments, and in experiments of short duration the uniformity of the results suggests that such a preparation may approach the normal state of the animal somewhat more closely than many other preparations that have been used. Until better instruments and methods are devised and used in conjunction with preparations which are capable of normal physiological responses, our knowledge concerning the normal and abnormal functioning of the coronary circulation will be necessarily limited as well as unavoidably inexact.

REFERENCES

(1) Spalteholz, W. Die Arterien der Herzwand. Leipzig, S. Herzel, 1924.
(2) Wearn, J. T. Chapter I. Diseases of the coronary arteries and cardiac pain. R. L. Levy, New York, Macmillan, 1936.
(3) Wearn, J. T. J. Exper. Med. **47**: 273, 1928.
(4) Shipley, R. A., L. J. Shipley and J. T. Wearn. J. Exper. Med. **65**: 29, 1937.
(5) Wearn, J. T. Bull. Johns Hopkins Hosp. **68**: 363, 1941.
(6) Wearn, J. T., S. R. Mettier, T. G. Klumpp and L. J. Zschiesche. Am. Heart J. **9**: 143, 1933–34.

(7) Gregg, D. E., R. E. Shipley and T. G. Bidder. Am. J. Physiol. **139**: 732, 1943.
(8) Grant, J. C. B. A method of anatomy. Williams & Wilkins, Baltimore, 1940, p. 507.
(9) Gregg, D. E. and R. E. Shipley. To be published.
(10) Hudson, C. L., A. R. Moritz and J. T. Wearn. J. Exper. Med. **56**: 919, 1932.
(11) Beck, C. S. Blood, heart and circulation. AAAS Pub. no. 13, 1940, p. 133.
(12) Katz, L. N., K. Jochim and A. Bohning. Am. J. Physiol. **122**: 236, 1938.
(13) Schmidt, C. F. Am. J. Physiol. **84**: 202, 1928.
(14) Gregg, D. E. and H. D. Green. Am. J. Physiol. **130**: 114, 1940.
(15) Johnson, J. R. and C. J. Wiggers. Am. J. Physiol. **118**: 38, 1937.
(16) Gregg, D. E., R. E. Shipley, R. W. Eckstein, A. Rotta and J. T. Wearn. Proc. Soc. Exper. Biol. and Med. **49**: 267, 1942.
(17) Shipley, R. E. and E. C. Crittenden, Jr. Proc. Soc. Exper. Biol. and Med. **56**: 103, 1944.
(18) Crittenden, E. C., Jr. and R. E. Shipley. Rev. Scient. Instruments **15**: 343, 1944.
(19) Soskin, S., W. S. Priest and W. J. Schultz. Am. J. Physiol. **108**: 107, 1934.
(20) Dumke, P. R. and C. F. Schmidt. Am. J. Physiol. **138**: 421, 1943.
(21) Bennett, S. H., W. H. Sweet and D. L. Basset. J. Clin. Investigation **23**: 200, 1944.
(22) Gregg, D. E. Blood, heart and circulation. Am. Assn. Adv. Sci. Publication no. 13, 1940, p. 81.
(23) Rössler, R. and W. Pascual. J. Physiol. **74**: 1, 1932.
(24) Green, H. D. and D. E. Gregg. Am. J. Physiol. **130**: 97, 1940
(25) Anrep, G V., J. C. Davis and E. Volhard. J. Physiol. **73**: 405, 1931.
(26) Stehle, R. L. and K. I. Melville. J. Pharmacol. and Exper. Therap. **46**: 471, 477, 1932.
(27) Bergmann, G. Ztschr. f. Biol. **98**: 536, 1938.
(28) Shipley, R. E., D. E. Gregg and E. F. Schroeder. Am. J. Physiol. **138**: 718, 1943.
(29) Kolin, A. Proc. Soc. Exper. Biol. and Med. **46**: 235, 1941.
(30) Katz, L. N. and K. Jochim. Medical physics. Year Book Publishers, Chicago, Illinois, 1944, p. 377.
(31) Rein, H. Ztschr. f. Biol. **87**: 394, 1928.
(32) Schmidt, C. F. and A. M. Walker. Proc. Soc. Exper. Biol. and Med. **33**: 346, 1935.
(33) Baldes, E. J. and J. F. Herrick. Proc. Soc. Exper. Biol. and Med. **37**: 432, 1937.
(34) Schmidt, C. F. and J. P. Hendrix. Proc. Assn. Res. Nerv. and Ment. Dis. **18**: 229, 1937.
(35) Gregg, D. E., W. H. Pritchard, R. W. Eckstein, R. E. Shipley, A. Rotta, J. Dingle, T. W. Steege and J. T. Wearn. Am. J. Physiol. **136**: 250, 1942.
(36) Shipley, R. E., D. E. Gregg and J. T. Wearn. Am. J. Physiol. **136**: 263, 1942.
(37) Pritchard, W. H., D. E. Gregg, R. E. Shipley and A. S. Weisberger. Am. J. Physiol. **138**: 731, 1943.
(38) Machella, T. E. Am. J. Physiol. **115**: 632, 1936.
(39) Gregg, D. E., H. D. Green and C. J. Wiggers. Am. J. Physiol. **112**: 362, 1935.
(40) Green, H. D., D. E. Gregg and C. J. Wiggers. Am. J. Physiol. **112**: 627, 1935.
(41) Green, H. D. and D. E. Gregg. Am. J. Physiol. **130**: 126, 1940.
(42) Gregg, D. E. and H. D. Green. Am. J. Physiol. **130**: 108, 1940.
(43) Anrep, G. V. and H. N. Segall. Heart **13**: 239, 1926.
(44) Peters, H. C. and M. B. Visscher. Am. Heart J. **11**: 273, 1936.
(45) Katz, L. N., W. Wise and K. Jochim. Am. J. Physiol. **143**: 479, 1945.
(46) Anrep, G. V. Lane Medical Lectures. Studies in cardiovascular regulation. Stanford University, California, 1936.
(47) Katz, L. N. and M. Mendlowitz. Am. J. Physiol. **122**: 262, 1936.
(48) Katz, L. N., K. Jochim and W. Weinstein. Am. J. Physiol. **122**: 252, 1938.
(49) Gregg, D. E. and D. Dewald. Am. J. Physiol. **124**: 435, 1938.
(50) Hamilton, W. F., I. Brotman and G. Brewer. Am. J. Physiol. **107**: 414, 1934.
(51) Johnson, J. R. and J. R. Di Palma. Am. J. Physiol. **125**: 234, 1939.

(52) JOHNSON, J. R. Fed. Proc. Am. Physiol. Soc. **4**: 37, 1945.
(53) GREGG, D. E. AND R. W. ECKSTEIN. Am. J. Physiol. **132**: 781, 1941.
(54) BOYER, N. H. AND H. D. GREEN. Am. Heart J. **21**: 199, 1941.
(55) GREEN, H. D., R. WEGRIA AND N. H. BOYER. J. Pharmacol. and Exper. Therap. **76**: 378, 1942.
(56) THORNTON, J. J. AND F. R. MAUTZ. Am. Heart J. **19**: 404, 1940.
(57) PHELPS, K. R. Am. Heart J. **21**: 664, 1941.
(58) SHIPLEY, R. E. AND D. E. GREGG. Am. J. Physiol. **141**: 289, 1944.
(59) ECKSTEIN, R. W., D. E. GREGG AND W. H. PRITCHARD. Am. J. Physiol. **132**: 351, 1941.
(60) GREGG, D. E., J. J. THORNTON AND F. R. MAUTZ. Am. J. Physiol. **127**: 161, 1939.
(61) MAUTZ, F. R. AND C. S. BECK. J. Thoracic Surg. **7**: 113, 1937.
(62) BECK, C. S. Ann. Surg. **118**: 788, 1943.
(63) BECK, C. S. AND A. E. MAKO. Am. Heart J. **21**: 767, 1941.
(64) GROSS, L., L. BLUM AND G. SILVERMAN. J. Exper. Med. **65**: 91, 1937.
(65) THORNTON, J. J. AND D. E. GREGG. Am. J. Physiol. **128**: 179, 1939.
(66) BURCHELL, H. B. Arch. Int. Med. **65**: 240, 1940.
(67) BURCHELL, H. B. Blood, heart and circulation. Am. Assn. Adv. Science, Publication no. 13, 1940, p. 139.
(68) ROBERTSON, H. F. Am. Heart J. **10**: 533, 1935.
(69) ROBERTSON, H. F. Surgery **9**: 1, 1941.
(70) MARKWALDER, J. AND E. H. STARLING. J. Physiol. **47**: 275, 1914.
(71) ANREP, G. V., A. BLALOCK AND M. HAMMOUDA. J. Physiol. **67**: 87, 1929.
(72) NONIDEZ, J. F. Am. J. Anat. **65**: 361, 1939.
(73) MOE, G. K. AND M. B. VISSCHER. Blood, heart and circulation. Am. Assn. Adv. Science, Publication no. 13, 1940, p. 100.
(74) LENDRUM, B., B. KONDO AND L. N. KATZ. Am. J. Physiol. **143**: 243, 1945.
(75) STELLA, G. J. Physiol. **73**: 36, 1931.
(76) LEARY, T. AND J. T. WEARN. Am. Heart J. **5**: 412, 1929–30.
(77) ECKSTEIN, R. W., J. T. ROBERTS, D. E. GREGG AND J. T. WEARN. Am. J. Physiol. **132**: 648, 1941.
(78) ANREP, G. V. AND H. HAUSLER. J. Physiol. **65**: 357, 1928.
(79) DAVIS, J. C. Am. J. Physiol. **123**: 50, 1938.
(80) VISSCHER, M. B. J. A. M. A. **113**: 987, 1939.
(81) GREGG, D. E. AND R. E. SHIPLEY. Am. J. Physiol. **142**: 44, 1944.
(82) GREGG, D. E., W. H. PRITCHARD, R. E. SHIPLEY AND J. T. WEARN. Am. J. Physiol. **139**: 726, 1943.
(83) GREGG, D. E. AND R. E. SHIPLEY. To be published.
(84) SHIPLEY, R. E., D. E. GREGG AND E. F. SCHROEDER. Am. J. Physiol. **138**: 718, 1943.
(85) GREGG, D. E. AND R. E. SHIPLEY. Am. J. Physiol. **141**; 382, 1944.
(86) GREEN, H. D. AND R. WEGRIA. Am. J. Physiol. **135**: 271, 1942.
(87) GREEN, H. D. Am. Assn. Adv. Sci. Publication no. 13, 1940, p. 105.
(88) BOYER, N. H. J. A. M. A. **122**: 306, 1943.
(89) SHIPLEY, R. E. AND K. G. KOHLSTAEDT. To be published.
(90) WOOLARD, H. H. J. Anat. **60**: 345, 1926A.
(91) WIGGERS, C. J. Chapter II. Diseases of the coronary arteries and cardiac pain. R. L. Levy, New York, Macmillan, 1936.
(92) MCDOWALL, R. J. S. Chapter VI. The control of the circulation of the blood. Longmans, Green, New York City, 1938.
(93) BRONK, D. W., L. K. FERGUSON, R. MARGARIA AND D. Y. SOLANDT. Am. J. Physiol. **117**: 237, 1936.
(94) BRONK, D. W., L. K. FERGUSON AND D. Y. SOLANDT. Proc. Soc. Exper. Biol. and Med. **31**: 579, 1934.
(95) SHIPLEY, R. E. AND D. E. GREGG. Am. J. Physiol. **143**: 396, 1945.
(96) KATZ, L. N. AND K. JOCHIM. Am. J. Physiol. **126**: 395, 1939.
(97) KATZ, L. N., K. JOCHIM AND A. BOHNING. Am. J. Physiol. **122**: 236, 1938.

RECENT EXPERIMENTAL STUDIES ON LEUKEMIA

J. FURTH

The Department of Pathology, Cornell University Medical College, New York Hospital, New York, N. Y.

Advancements in knowledge of virus disease and of cancer had thoroughly prepared the ground for leukemia research so that during the third decade of this century investigators in different parts of the world discovered independently that leukemia of mammals is transmissible. The causation of chicken leukosis[1] by a filterable agent had already been known (24) and studies of avian leukosis disclosed numerous facts on the nature of the agents causing leukosis which proved to be variants of chicken tumor viruses. However, as knowledge advanced a gap appeared between avian and mammalian leukemias and it seemed that the solution of major problems of human leukemia would come from studies of this disease in mammals in which leukemia appears to be neither induced nor transmitted by virus. Recent research with plants, protozoa and animals disclosed the existence of transmissible cytoplasmic agents determining the character of cells. Physiological and pathological changes in cells are now viewed together as alterations of the same intrinsic cellular mechanisms which determine the character of the cell (19, 47). Avian leukosis, which readily yields a cytoplasmic agent modifying cells, may yet become the object of fundamental research.

The basic change in leukemia is an alteration in the cell with acquisition of properties of a cancer cell including partial or complete independence from factors regulating growth and differentiation of normal cells. The ease of induction of this change (leukemogenesis), the development of relatively homozygous strains of mice in which this change seems inevitable, the isolation of lines of transmissible leukemic cells of different types, have laid the foundations of modern research on mammalian leukemia.

Historical aspects of these developments and their earlier phases have been well reviewed (25, 86, 32, 96, 38, 33b, 88, 60b, 17), and this review surveys only researches of the last few years. To introduce these, it seems desirable to recapitulate conclusions of earlier observations.

RÉSUMÉ OF EARLIER WORK. Leukemia of mammals can be transmitted by malignant (leukemic) blood cells which proliferate in any genetically compatible new host, retaining their essential character, and bringing about the death of the host.

The type of leukemia is determined by the character of the leukemic cell, the site of malignant change and the resistance of the host. The same line of leukemic cells may form tumor-like masses (lymphosarcoma or myelosarcoma) or systemic disease with no tumors, depending on the portal of entry (site of origin); the degrees of ability to form tumors or invade the blood stream are inherent characteristics of malignant blood cells.

[1] Term proposed by Ellermann to include all types of leukemia.

Strains of mice have been developed by continued brother and sister matings in which the incidence of leukemia is fairly constant (high, medium or low) and is characteristic of the strain. This incidence is determined by genetical factors, modified by such nongenetical influences as acquired through the foster mother, caloric intake, etc. If powerful, hereditary or extrinsic factors alone seem capable of producing the disease.

Avian leukemia does not behave as a contagious disease yet it can be transmitted by cell free agents. Some agents produce leukemia, others sarcoma and still others may produce both. Myelo- and erythroleukosis are easily transmissible and one agent usually produces both. Only a few cases of lymphoid leukosis, which is one of the most common diseases in chickens, proved transmissible. Under usual conditions of laboratory passages the range of activity of an agent does not change.

What are the hereditary factors, what is the magnitude and character of maternal, hormonal, nutritional and other influences which determine the incidence of leukemia, what is the change that takes place when a normal cell turns leukemic, and given such cells what are the forces that influence their growth, were the major topics of recent investigations on leukemia of mammals. What is the nature of the agent of avian leukosis, the cause of the common avian lymphomatosis associated with paralysis and to what extent can a virus be changed, were the major problems of avian leukosis recently investigated.

Nomenclature. The term leukemia or leukosis implies neoplasia of a blood cell (hemoblastosis) including both the aleukemic and leukemic types of the disease. Lymphosarcoma is used to designate tumors formed by lymphocytes; myelosarcoma for those of myelocytes; these are regarded as types of leukemia. The suggestion to name leukemic blood cells malignant lymphocytes or myelocytes found wide acceptance. It should be considered whether a prefix "mal" for malignant or "neo" for neoplastic would not simplify nomenclature (e.g., neolymphocyte). Leukemogenesis is the process resulting in the development of leukemia. Leukemogens are agents producing the leukemic cell. Bioassay for leukemia is the determination of the presence of malignant blood cells by the injection of the cell suspension into animals highly susceptible to the cells to be assayed, at an age when they are free from leukemia.

THE GENESIS OF LEUKEMIA. Many factors co-operate in the causation of neoplasms. The ever present modifying factor is the genetic constitution of the host. Nursing conditions, hormonal activity, nutritional state and other factors inherent to life are leukemogenic influences of varying intensity. Furthermore, organisms are subject to accidental exposure to chemical, physical and other agents which may be powerful leukemogens. These factors are interrelated; e.g., different agents can cause the same nutritional deficiency. Vitamins may affect neoplasms through indirect effect on endocrine organs (12a); one agent may enhance or depress the action of another.

The study of these influences requires the use of several inbred strains of animals because of genetic differences in response to given stimuli; e.g., the C3H mice are susceptible to the induction of leukemia with estrogens but not

to that with methylcholanthrene and the reverse is true for mice of the F strain (62d).

Heredity. The familial occurrence of leukemia in animals has been shown to be due to genetic factors. By continued brother and sister matings relatively homozygous strains of mice have been developed in which a high percentage of the animals die of spontaneous leukemia. Non-leukemic males and females of such strains, mated with each other, yield mice with the same liability to leukemia as leukemic males and females, indicating the significance of non-hereditary factors even in mice with hereditary liability. The fact that a certain number of mice of a highly inbred strain fails to develop leukemia may be explained in the opinion of R. R. Gates (40) by the conceptions of penetrance or by Heisenberg's principle of indeterminancy. An illustration of the latter is radium decay which proceeds at a fixed rate regardless of extrinsic conditions, although it cannot be predicted which atoms will break down at a given time. So, in homozygous strains, only a certain number of mice will develop leukemia at a given age even if environmental conditions are identical (40).

Such high leukemia strains of mice are *a*, C57 of MacDowell with an expression of leukemia of about 90 per cent; *b*, our Ak strain with about 70 per cent, and *c*, Strong's F strain (62a) with about 50 per cent of leukemia. These figures are approximate. There is some variation in different batches of the same strain due to non-genetic modifying factors (80a, 81b, 18). In our Ak mice the figures vary from about 60 to 80 per cent; they are usually low when intercurrent diseases are prevalent and high when they are absent.

In hybrid mice between high and low leukemia strains not only is the incidence of leukemia lower than in the high leukemia strain but the onset of the illness is markedly delayed.

In mice of the high leukemia strain spontaneous leukemia usually occurs when the mice are about 6 to 9 months of age. When leukemic cells are introduced into immature mice they produce leukemia rapidly; hence, young animals are excellent objects for the identification of malignant blood cells, a test termed bioassay. F1 hybrids between high and low leukemia strains are usually susceptible to grafts of leukemic cells arising in their parents even though they are but slightly or moderately liable to develop spontaneous leukemia. Bioassays indicate that the leukemic cells are not present at birth in normal mice of a genetically susceptible strain and that the malignant change precedes shortly the onset of leukemia (84a, 98).

Cohnheim's theory that neoplasms originate in cells misplaced during embryonal life or in cells remaining in their normal sites but so constituted as to become malignant after a given age, or in cells that are perhaps malignant at birth but are restrained until conditions favor their multiplication, has been revived by Fischer. He investigated the genesis of carcinoma of the breast, a neoplasm that arises in an organ undeveloped at immature age. The sequence of changes is particularly conspicuous in male animals; first, the development of breast tissue followed by hyperplasia, which under intensified physiological stimulation may go on to neoplasia at multiple sites, particularly when the "milk agent"

is carried by the mouse. Thus cancer may arise in cells that are non-existent before the stimulus producing it is applied.

The tools of genetic studies are the inbred strains. The mode of inheritance of leukemia varies with different strains and the negative influence from the low leukemia strain is as important as the positive influence from the high leukemia strain. A high leukemia strain crossed with one low leukemia strain (Ak x Rf) may yield few leukemias while the same high leukemia strain crossed with another low leukemia strain (Ak x C3H) will yield many leukemias (37). The genetic influence of the low leukemia strain Rf is so strong that in the first back cross generation to this strain the incidence of leukemia is almost as low as in the low leukemia strain. The incidence of leukemia in F1 hybrids of the F strain with several low leukemia strains is about as high as in the F strain (60).

Mercier crossed a high leukemia strain (with an expression of leukemia of 49.2 per cent) to a low leukemia strain and none of the F1 hybrids developed leukemia, leading to a conclusion identical with that of Slye (103), that leukemia is inherited as a recessive character. In the F2 hybrids leukemia reappeared. However it is possible that the same high leukemia strain crossed to another low leukemia strain (e.g., C3H) would yield a high percentage of leukemic mice. Indeed, it is possible that the hybrids of two low leukemia strains will yield many leukemias. What matters is the combination of genes, and what this will yield in terms of leukemia only a test can tell.

Surveying the reported data on two sets of crosses MacDowell et al. conclude that the proportion of leukemias is a mathematical function of the total heredity from the high leukemia strain. In crosses C58 x StoLi, the function appeared to be directly arithmetical (81b). In crosses of Ak x Rf the function was semi-logarithmic with the logarithm of the percentage of leukemias varying with the proportion of high leukemia heredity (18). This difference may, in their opinion, be due to extrinsic variables and with the number of such variables unknown such experiments do not define the number of genes responsible for the parental influence (81b). The mere appearance of a neoplasm in an animal or the frequency of neoplasms in an inbred strain is no indication of intrinsic tendencies. The unique characteristic of genes is their assortment in cross breedings (82).

The complexity of the genetic basis of susceptibility to leukemia is illustrated by the second back cross test of MacDowell et al. (81b) for determiners of spontaneous leukemia. In the F1 generation, theoretically all mice are genetically alike but in the back cross generation the number of different genotypes increases with the number of genes involved in the character studied. If one or two major genes were involved in susceptibility to spontaneous leukemia, then there would be only two or four different genotypes with respect to leukemia among the males from the first backcross generation. The data indicate that the males of the first backcross generation were of many genotypes and this is in agreement with the hypothesis that spontaneous leukemia depends for its development upon many genetic factors. The differences in the results of various workers can be explained by genic differences among the strains used

and by the many unknown extrinsic variables. Environmental factors modify the expression of genetic tendencies and different types of leukemia may have different genetic bases (18).

An admirable analysis of the numerous variables: mother's age, strain of nurse, sex, hair color and longevity in the same mating combinations, is contained in the recent paper of MacDowell et al. (81b). Longevity itself is influenced by the genetic constitution of the father, by the parturition age of the mother, by the strain of the nursing female and by the sex, each influence acting differently (81b).

The evidence that genes are a strong influence on the incidence of leukemia is conclusive but it is unknown how many genes are involved and how they exert this effect. Experiments to be described below suggest that the leukemic cell is not inherited as such and that the malignant transformation takes place in certain loci during life. If one considers that differences in response to leukemogenic agents often do not parallel differences in liability to spontaneous leukemia and that susceptibility to one agent does not indicate susceptibility to another agent it becomes clear that we are far from knowing the rôle of heredity to spontaneous leukemia.

Data available on the heredity to leukemia in man are meagre and inconclusive. The heterozygocity of the human population and strong influence from non-leukemic parents may mask genetic tendencies. The occasional appearance of this disease in infancy certainly indicates a congenital if not a genetical factor.

A maternal influence in the occurrence of leukemia is indicated by the incidence of leukemia in reciprocal matings.

	MATING COMBINATION H/L*		
	C58/StoLi	Ak/Rf	Ak/C3H
	Author		
	MacDowell et al.	Cole et al.	Furth et al.
High leukemia strain	89.6	69.3	58
H♀ x L ♂.....	61.9	21.9	50
L ♀ x H ♂.....	42.5	11.6	34

*H = high leukemia strain.
L = low leukemia strain.

The existence of a maternal influence led to a search for a milk factor similar to that of Bittner's milk agent. Foster nursing by low leukemia dams reduces the incidence of leukemia in the high leukemia strain to about one-half but mice of a low leukemia strain nursed with high leukemia strain dams do not develop leukemia. This lowering of the incidence of leukemia cannot be explained by failure of the mice to reach the leukemia age (2, 33c) since mice fostered by low leukemia strain mothers live on the average longer than those fostered by their own mothers. If C3H dams, which carry a strong milk factor,

are used for foster nursing high leukemia strain mice, the latter will develop mammary gland carcinoma while the C3H mice will not develop leukemia. The incidence of leukemia in F mice fostered by low leukemia strain mice was lower (28 per cent at 500 days) than in unfostered mice (38 per cent at 500 days); the reciprocal foster nursing did not produce leukemia in the low leukemia strain mice (62c).

Thus nursing influence for leukemia, unlike that for mammary gland carcinoma, is not transmitted to offspring yet it may be related to the virus-like milk factor for mammary gland carcinoma. If this agent is conceived to be a virus, it is subject to reactive forces of the host which may mask its existence as is the case with the virus of Shope (cf. Rous).

The superiority of a strain ("B alb") in nursing has been demonstrated by MacDowell et al. (cf. 81b) in terms of body weight, tail length and lengthening of life. This effect is greater in males and is gradually lost during nine months. In MacDowell's experiment (81b) the nurse strain influence appears to be due primarily to the accidental presence of certain sex limited non-leukemic conditions shortening life. This does not seem likely in our experiments (33c).

The parturition age of the mother is a strong modifier of genetic factors. The increasing parturition age of the mother progressively delays the appearance of leukemia (81b). Thus, youngest mothers reveal more accurately the potential leukemics. The nature of this phenomenon is not known. Chemical and physiological changes occur with increasing age. The influence might be imparted to the young through the egg cytoplasm or through the placental circulation, or through the first milk (81b). It may be related to the nursing influence discussed. The placenta is no barrier to maternal hormones or to agents of the size of viruses—though the milk "virus" does not seem to pass through the placenta. Immunization of the mother against substances in the fetus is intensified during subsequent pregnancies as indicated by observation on the Rh factor. Is a similar neutralization phenomenon operative against the hypothetical agent producing leukemia or do the factors causing leukemia begin to operate during embryonal life and decline in intensity with advancing age of the mother?

The observations described indicate that susceptibility to leukemia is at its peak when the high leukemia strain mouse is nursed by its mother and she is young when she nurses.

Sex and sex hormones. A sex factor is operative in most strains of mice. Data tabulated by Cole (18) show that in most strains females have a higher incidence of leukemia than males. It is remarkable that in man the reverse is true, the ratio being about 2:1 in favor of males.

This sex difference is small in mice, and may not be statistically significant, as in the recent experiments of MacDowell et al. (81b). These workers explain the higher incidence of leukemia among females in their experiments by the greater longevity of the females. They suppose that females resist certain factors that shorten the life of males and in so doing permit more of the potential leukemics to develop leukemia (81b). Mortality curves in other strains indi-

cate however that most males survive their leukemic sisters. In view of the unquestionable leukemogenic effects of estrogens, the greater incidence of leukemia among the females should be considered to be due in part at least to the female sex hormone. But thus far, experiments made to verify this supposition have not yielded uniform results.

In the high leukemia strain (Ak) the incidence of this disease was lowered from 74 per cent to 45 per cent by ovariectomy performed at 23 to 56 days of age. Among males subjected to orchidectomy at 20 to 56 days the incidence was 60 per cent as compared with 52 per cent among the controls (84). In Gardner's strains in which the incidence of leukemia in females is not higher than in males, the ovariectomy had no significant effect (41). Similarly in the F strain of Strong (60a) the incidence of leukemia is about the same in both sexes and gonadectomy did not significantly alter these figures.

In a strain of Lynch (R.I.L.) the incidence of leukemia in the females was 88.4 per cent, in males 53.5 per cent. Castration in males raised this figure to 97 per cent while the administration of testosterone propionate to castrates lowered the latter figure to 58.3 per cent. Ovariectomy had no significant effect on the incidence of leukemia (78b). Removal of the gonads in these experiments was performed at about weaning age or shortly thereafter. Murphy suggests that the sex difference in susceptibility to spontaneous leukemia is due to an inhibitory effect by the male sex hormone rather than to a stimulation by the ovarian secretion. But the leukemogenic effect of estrogenic hormones is amply proven by the experiments of others. The interpretation of all of these experiments is rendered difficult by the unknown degree of compensation of the adrenals for the loss of ovarian sex hormones. In man, relapses after castration at old age for carcinoma of the prostate are attributed to the compensatory production of the male sex hormone. Castration at one day of age leads to hyperplasia of adrenals and occasional tumor formation with excessive sex hormone production. It would be of interest to know the effect of castration at various ages on the incidence of leukemia and the alteration of the hormonal balance by castration.

The leukemogenic action of sex hormones is indicated by the experiments of Gardner et al. (41, 42), Lacassagne and Bischoff et al. and of others (cf. 42). The well controlled experiments of Gardner indicate that in certain strains estrogens are as powerful leukemogens as any known agent (41). Other strains of mice (e.g., C57 of Gardner) are entirely refractory to the leukemogenic action of estrogens even though other organs of the same mice are highly responsive to them. The leukemogenic action of the chemicals tested roughly parallels their estrogenic power and can be neutralized by androgenic hormone. Testosterone does not influence the occurrence of leukemia in a low leukemia strain of mice (41). Discontinuation of estrogens after $2\frac{1}{2}$ to 3 months resulted in the production of more lymphoid neoplasms than continuous exposure, presumably because the animals lived longer. Thus, a short but intense treatment is capable of bringing about irreversible changes leading to leukemia. It will be of great interest to learn how brief and how intense this treatment has to be to produce this change when other conditions are set for it.

Susceptibility to estrogens is under genetic control of the host (41). In hybrid mice with known susceptibility to estrogens the leukemogenic action of estrogens varied greatly. Hybrid mice born of parents of two susceptible strains had the highest incidence of leukemia following estrogen treatment, none occurring in the untreated controls. When both parents were of the resistant strain the hybrids had in one combination a greater, and in another combination a lesser susceptibility to estrogens than the parents (41).

How do estrogens produce leukemia? Their primary effect is an atrophy of the lymphoid organs including the thymus and bony proliferation with replacement of the marrow. The leukemias produced are of the lymphoid type and arise in the atrophic organs, frequently in the mediastinum, and are indistinguishable from spontaneous leukemias of the same type. Mediation by the adrenals is not required. Estrogens act on lymphoid organs even in the absence of adrenals (101). X-rays are also leukemogenic and the primary change is likewise that of depression of hemopoiesis. Neoplasms often arise in atrophic organs, but the factors which push focal areas of regeneration over to the neoplastic side remain a mystery.

Thymus. It has been noted that the thymus is a frequent primary site of leukemia in a strain of mice (Ak) with high hereditary liability to spontaneous leukemia. Removal of this organ in this strain at 31 to 71 days of age resulted in a reduction of the incidence of spontaneous leukemia from 77 to 8 per cent in females and from 61 to 11 per cent in males (84). Removal of the spleen at comparable ages had no effect on the incidence of spontaneous leukemia. The life span of thymectomized Ak mice is on the average about one half longer than in normal Ak mice and this is probably due to prevention of death from leukemia.

The presence of the thymus is important in leukemogenesis in at least this strain of mice. Moderate doses of methylcholanthrene produce leukemia sooner and in larger numbers of animals whose thymus is present (33e). In this strain the thymus is frequently the primary site of the disease and it may be supposed that its lymphoid cells are especially liable to a malignant change.

Investigations on the possible mechanisms by which the incidence of leukemia is lowered by thymectomy disclosed the following: *a*, the effect of thymectomy is not inherited, the incidence of leukemia being as high in the offspring of thymectomized mice as in their normal grandparents (33e); *b*, thymectomized mice are as susceptible to the growth of grafted leukemic cells as normal mice (33e); *c*, removal of the thymus reduces susceptibility to the leukemogenic action of methylcholanthrene but not to the carcinogenic action of this hydrocarbon; *d*, administration of thymus from mice of a low leukemia strain to thymectomized mice over a period of four months does not seem to restore susceptibility to leukemia (33e). However, it is possible that the "substitution therapy" was not adequate.

These observations suggest that in mice of the high leukemia strain Ak, potentially neoplastic blood cells reside in the thymus and that removal of this organ causes reduction in the incidence of leukemia not by virtue of an endocrine factor but by removal of potentially malignant cells.

If the above assumption is correct, it may be expected that the Ak thymuses retain malignant potentialities after transplantation in a manner as limbs of creeper fowl embryos develop as achondroplastic when grafted on normal embryos. It seems desirable to remove the thymus in strains of mice in which leukemia is myeloid or lymphoid but not thymic in origin and to study the effect of thymectomy on other neoplasms.

Accidental involution of the thymus often accompanies malnutrition, intoxication or disease. The experiments described show how these can modify an expression of genetic susceptibility to lymphoid leukemia.

Adrenal. The reciprocal relationship between thymus and adrenal cortex is well known (cf 89, 22a, 22b). It follows that hyperplasia, hypoplasia or dysfunction of the adrenal cortex may modify the incidence of spontaneous leukemia or the response to a leukemogen. However, there are no experimental studies on record on such an effect. The conspicuous effect of adrenal cortical hormones on the growth of leukemic cells is discussed below.

Nutritional influence. Tannenbaum (108a, 108b) has presented statistical and experimental evidence indicating that restriction of caloric intake reduces the incidence of neoplasms, in man and animals. Weight loss alone, what ever its cause, can be regarded as exerting an inhibiting influence on the incidence of neoplastic diseases.

The effect of reduction in food intake on the genesis of spontaneous leukemia is indicated by the experiments of Saxton et al. (98). In mice of a high leukemia strain receiving limited amounts of a qualitatively adequate diet, the incidence of leukemia was 10 per cent as compared to 65 per cent in normally fed controls. Bioassays at 9 to 12 months indicated the absence of malignant lymphocytes in the underfed mice, and disclosed their presence at 14, 16 and 17 months. In normally fed mice of the same strain, malignant cells appeared at 7 to 13 months. These experiments indicate that *a*, malignant lymphoid cells are absent in young normal mice of the high leukemia strain Ak; *b*, the malignant transformation takes place shortly before leukemia becomes manifest; and *c*, this transformation is delayed but not entirely prevented by underfeeding (98).

It is possible that the effects here described are due to some specific nutritional deficiency due to the overall reduction of food intake although none was noted. Underdevelopment of thymus and lymphoid tissue might possibly explain a lowering of the incidence of leukemias but they can hardly account for the inhibition of other neoplasms by underfeeding. Thymectomy produces a marked reduction in the incidence of leukemia without causing weight loss of the experimental animals (33e).

In mice "painted" with methylcholanthrene, a 50 per cent caloric restriction reduced the incidence of leukemia from 96.2 per cent to 35 per cent and increased the mean latent period from 73 to 237 days (115).

Lack of cystine in the diet also reduced the incidence of leukemia in mice "painted" with methylcholanthrene. The addition of gelatine to a low cystine diet raised the incidence of induced leukemia from 10 to 32 per cent and this effect is explained by the presence of a small amount of cystine in gelatine (116).

There are no data on the effect of low cystine diet on the incidence of spontaneous leukemia.

Nutritional deficiency has been known to cause involution of the thymus. Deficiency of riboflavin, pantothenic acid and thiamin or pyridoxin can cause depression of the weight of the thymus (106). Depression of lymphoid organs usually accompanies atrophy of the thymus and any agent which brings about atrophy of this organ might influence the incidence of lymphoid leukemia.

Chemical leukemogenesis. Earlier experiments suggested that tar and carcinogenic hydrocarbons are also leukemogenic (10, 25). Morton and Mider described a readily reproducible method to induce leukemia with such an agent. They "painted" repeatedly a solution of methylcholanthrene on the skin of mice at different sites so as to minimize the development of local tumors. This procedure, they believed, hastened the onset of leukemia in genetically susceptible strains of mice. However, some low leukemia strains proved equally susceptible to this hydrocarbon.

The quantity of methylcholanthrene sufficient to produce leukemia is small. Six "paintings" with a 0.5 per cent solution of methylcholanthrene in benzol applied twice weekly over an area of a few centimeters has raised the leukemia incidence and hastened its onset. Two paintings had little effect; eight paintings were better than six but more intense treatment did not raise the leukemia incidence (33d).

The efficiency of a hydrocarbon in hastening the appearance of leukemia in the high leukemia strain F bears a direct relation to the potency of the hydrocarbon to induce breast tumors (62d).

There are differences in susceptibility to leukemogens among different strains of mice (61). A strain resistant to the leukemogenic action of x-rays may be susceptible to that of methylcholanthrene and vice versa (60b).

The leukemias produced by hydrocarbons are usually of the lymphoid type and "atypical" (see below). There is no clear demonstration of the ability of these substances to induce myeloid and monocytic leukemias although there are experiments on record suggesting such potentialities (33b).

In most of these experiments benzol has been used as solvent. Benzol itself is leukemogenic (cf. 60b) but not under the conditions tested. Solvents may have a pro or anti-carcinogenic action, though alone they do not produce new growth (12b).

The leukemogenic action of benzol is suggested by human observations (32); the few experimental studies on record suggest that it is a weak leukemogen (cf. 25, 60b). The work of Büngeler et al. (60b) suggesting that indol is leukemogenic remained unconfirmed.

2-amino and 2 acetoaminofluorene when fed to rats produces neoplasms in different organs (118) including leukemias. Bielschowsky (a) noted four leukemias in 104 rats of a non-leukemic strain fed on this compound but most of these rats had other neoplasms affecting the liver, mammary gland, etc., and these may have masked the leukemogenic potentialities of this compound.

Later studies of Bielschowsky (b) suggest that as a coleukemogen this compound might be unusually effective.

Physical leukemogens. X-rays have long been known to cause leukemia in man (cf. 32, 52) and they produce it readily in mice. According to a recent survey the incidence of leukemia in radiologists for a fifteen year period is ten times greater than in physicians who are not radiologists (70).

There is little known about the factors of irradiation responsible for the development of leukemia. The doses which have this effect are large enough to destroy the bone marrow. Regeneration follows atrophy but the stigmata of massive irradiation persist in experimental animals. Hemopoietic tissues once massively irradiated may never be normal. The spleen shrinks, connective tissue increases in amount and the organ is laden with hemosiderin filled macrophages. Large areas of atrophy or necrosis have been a common autopsy finding in the femoral marrow of old mice that had been given about 350r at approximately six weeks of age (33d). Some cells among those newly formed acquire the property of unrestrained growth. Both myeloid and lymphoid leukemias are produced by x-rays. Henshaw (b) has been most successful in inducing leukemia with x-rays. He has given doses of 200r for 3 to 4 month old mice, repeated at four week intervals for twenty weeks (total of 1200r) and thereby produced leukemia in 30 per cent of C57 mice. Blood counts showed in general, leukopenia, but in some animals, recovery was followed by marked leukocytosis with leukemic infiltration at death. Leukemia was predominantly of the primitive cell type and death occurred usually between the seventh and twelfth months. However, the incidence of leukemia among his controls was also high (7 per cent). There is no study on record which would give quantitative information on the relation of the dose and quality of x-rays and mode of application, to the incidence of leukemia. Neither are the meager data on strain differences in susceptibility to x-rays convincing. It appears from the survey of the literature that x-rays are weak leukemogens and act after a long incubation period. In a small series, irradiation of F1 hybrids between a high and low leukemia strain (Ak/Rf) at 4 to 6 weeks of age, a single dose of 87r had no definite leukemogenic effect, though this dose was productive of ovarian neoplasms of the granulosa cell type; 175r was slightly and 350r markedly leukemogenic (33d).

There is no knowledge on the leukemogenic capacities of other radiating energy. Their action is similar to that of x-rays and they might also have a leukemogenic power.

Coleukemogenesis. Leukemogenic agents to which man is exposed are numerous and to understand the genesis of spontaneous leukemia it is of great interest to find out if different agents which alone are not or are only slightly leukemogenic enhance or retard the action of leukemogens. X-rays proved to be powerful co-carcinogens (77, 72), and the following considerations led us to test their effect on leukemogenesis. After the application of x-rays destruction of blood forming organs is followed by regeneration during which mitotic figures are seen in abundance. It seemed probable that during this phase the blood forming organs

would be particularly susceptible to another leukemogen. The experiments performed indicate that pre-irradiation with 175r enhances the susceptibility of mice to the leukemogenic action of small doses of methylcholanthrene, the leukemias appearing earlier and in larger numbers than in mice subjected to these agents alone (35a). Current experiments have shown that 87r applied under similar conditions are less and 350r are more effective than 175r in enhancing susceptibility to methylcholanthrene.

The combined effect of x-rays and methylcholanthrene was tested by Kaplan and Kirschbaum under entirely different conditions. X-rays were given in doses of 80r until a total of 720r and 880r were given. The application of methylcholanthrene was so intense that 80 per cent of the mice died of leukemia at an average of 114 days as contrasted with 58 per cent of mice exposed to both agents, dying at an average of 127 days. The authors are inclined to explain the apparent discrepancy between their results and ours (35a) by differences in the genetic constitution of the strain of mice used, but they also emphasize the higher dosage and different method of irradiation in their experiment. As Tannenbaum points out, cocarcinogenic and anticarcinogenic actions are usually effective only during one stage of carcinogenesis (108c). The procedures of the two groups of investigators are not comparable and these may account entirely for the differences in the results.

There are other leukemogens whose co-operative potencies deserve further study; e.g., estrogens are known to be leukemogenic and carcinogenic in large quantities, and procarcinogenic (43) in small quantities. Among the leukemogenic substances the pluripotent acetoaminofluorene has a remarkable carcinogenic action on stimulated cells. It readily produces carcinoma of the thyroid gland that had been rendered hyperplastic with thiouracil (7b). Any agent that stimulates hemopoietic organs may prepare the ground for a leukemogen, and may cause neoplasia of blood cells when small doses of a leukemogen also reach hemopoietic foci. Hereditary susceptibility to leukemia may be localized to one lymphoid organ as suggested by thymectomy experiments (84). However if an organism is exposed to large quantities of a carcinogenic agent, neoplastic cells probably arise in many foci; e.g., methylcholanthrene produced innumerable discrete lung tumors in almost every mouse that had been painted with leukemogenic doses of this hydrocarbon.

Urinary substances stimulating leukopoiesis. The idea that leukemia may result from disturbance of the unknown mechanism which maintains leukocyte level will now be considered. Stimuli producing transient depressions and elevations of this level are well known and it is conceivable that a disturbance of this mechanism similar to that of an endocrine gland might result in a lasting rise of the leukocyte level and in an accompanying hyperplasia of a hemopoietic system. Thus, a leukemia-like change might result but this supposition does not explain the change in the immature blood cells endowing them with unrestrained growth in new hosts into which they are grafted.

Wearn et al. (cf. 53, 50) discovered that the urine of leukemic patients contains substances that stimulate hemopoiesis in guinea pigs. Miller and Turner

(75, 110) believe that these substances are specific stimulators of hemopoietic systems. All of 36 animals given this principle from patients with chronic myeloid leukemia are stated to have developed changes similar to that of human myeloid leukemia. Their figures 1 to 3 illustrating these changes show, in the reviewer's opinion, hemopoiesis not definitely leukemic in character. All of 40 animals given the principle from patients with chronic lymphoid leukemia are described as having developed comparable changes. In 6 guinea pigs given the material from patients with Hodgkin's disease, a "trend toward lymphoid stimulation was apparent with the character of lymphosarcoma in some animals." The material with lymphoid activity is believed to be a hydroxy-acid; that with myeloid activity a non-carbinol acid. Patients with monocytic leukemia and Hodgkin's disease are excreting both in large quantities but one substance can be converted into the other. Normal people are assumed to excrete similar substances in small quantities although none has been demonstrated. The most potent substance that produces myeloid hyperplasia and metaplasia in guinea pigs is in the urine of patients with myeloid leukemia. The active principle can be concentrated by the benzoic acid absorption technic. A different substance with similar activity obtained by Hirschman et al. seems to be a protein or glycoprotein.

No attempt is described by these investigators to prove transmissibility of the disease by cells. Leukemia of the guinea pig is known to be transmissible (25) as are most, if not all, leukemias of mice. The effect of similar extracts from the urine of patients with other neoplasms has not been adequately tested. In mice, advanced non-malignant extramedullary hemopoiesis frequently accompanies tumors of different sorts (65, 4). The animals of Miller and Turner usually became moribund at 3 to 5 weeks after the beginning of the treatment—an unlikely brevity and course for a neoplasm.

These authors interpret their data as supporting the well known "theory" of Ziegler, which postulates a balance between myelcid and lymphoid tissues and hyperplasia of one system consequent to hypoplasia of the other but until better evidence is presented this "theory" will remain an assumption. Wearn and his present associates regard the changes described as leukemoid but such changes are non-specific. The chemical and biological properties of substances producing them and their relation to lymphoid and myeloid leukemias deserve further study.

Leukemoid reactions. The hemopoietic reserve of the bone marrow appears less in rodents than in man, and it is common to note orderly myelo- and erythropoiesis in the liver and spleen of older rodents and during the course of varied infections and intoxications. Most of the greatly enlarged spleen and less often as much as one-third of the enlarged liver may be hemopoietic tissue and the leukocyte count may exceed 200,000. Indeed, in extent, these hemopoietic foci may equal that encountered in the usual cases of leukemia. These hemopoietic cells are, however, not autonomous and they cannot be grafted on normal related mice. They also lack the invasive character of malignant blood cells. Non-malignant extramedullary hemopoiesis (4, 65, 90, 33d)

often accompanies chemical carcinogenesis and grafted tumors of different sorts, particularly when the latter are large and undergo necrosis or secondary inflammation. With removal of the grafted tumor and healing of wounds, or cessation of causative infection, they regress. The morphological criteria differentiating it from leukemias have been well described by Barnes and Sisman. Occasionally it is not possible to distinguish between these two fundamentally different processes by microscopic examination. Numerous bioassays were performed to exclude the leukemic character of this extramedullary hemopoiesis and with rare exception the inoculation into young mice of the homologous strain failed to yield leukemia (33d). It is possible that this advanced hyperplastic change is mildly preleukemic.

The leukemic cell. Almost all investigators of experimental leukemia arrived at the conclusion that the essential change in leukemia is a malignant transformation of an immature blood cell and that a single locus of lymphoid neoplasia can produce as an end result either leukemia or lymphosarcoma, depending on the character of the malignant lymphocytes. The factors responsible for tissue invasiveness and degree of blood involvement are unknown (33b, 60b). It is possible to produce tumor-like masses in the subcutaneous tissue by the introduction of malignant blood cells in the subcutaneous tissue even from cases with diffuse involvement, formerly spoken of as systemic disease, not accompanied by tumor formation (33d).

The idea that the basic change in leukemia is one of interference with maturation of primitive cells with consequent proliferation lacks experimental support. In Potter's opinion the malignant transformation is "the result of a depression in the rate of differentiation which accumulates cells" (cf. 83). But in some myeloid leukemias there is a conspicuous maturation and this view does not explain the essential alteration in leukemia which takes place in the cell.

The idea that lymphocytes with leukemic properties in one animal may behave as normal cells on transplantation into an animal with a different genetic constitution (80b) has no basis. Experimental data indicate that leukemic cells under such conditions do not proliferate and the studies of Loeb imply that they are likely to be destroyed by the inflammatory reaction provoked by the genetic differential between host and recipient. There are no data on record that suggest that leukemic cells re-acquire the basic properties of a normal cell under any condition. If such a process should occur it would indicate a possible means of controlling neoplastic growths.

The experimental observations indicate that the leukemic cell is a newly formed cell with individual properties of its own often perplex in terms of normal maturation; e.g., in a malignant myeloid cell (as is common in myeloid leukemias of chickens) the nuclear lobation may be advanced, yet granulation is scanty if any. One type of leukemic cell in mice is characterized by exaggerated nuclear lobations, absence of cytoplasmic granules and of oxydase enzymes and lack of phagocytic ability. This was first described as an "atypical leukemia" (3a) but subsequent observations have shown this to be a fairly characteristic type although its genesis and character is subject to controversy (66). Occa-

sionally the typing of leukemias in man and mice offers considerable difficulty. Organ affinities are of little help in their identification since any blood forming organ can be the favored site of any type of malignant blood cell. In transmission experiments the three types of blood cells (lymphocytes, granulocytes and monocytes) remain true to type and transitions or transformations have not been observed. These observations support the idea of independence of these cell types in the adult animal.

Neoplasms arising in reticular or endothelial cells of hemopoietic organs have also been observed in rats (29) and in mice and transmitted in series (3b, 39). These cells remain true to type and fail to produce any type of blood cell. One of these could be readily grafted on the chorioallantois of the chick embryo producing a bulky growth in eight days. There seems to be a relation between neoplasm of monocytes or reticular cells and Hodgkin's disease (33a). A cytoplasmic inclusion body in lymph nodes characteristic of Hodgkin's disease described by Grand deserves further attention.

It is beyond the scope of the present review to survey in detail chemical, metabolic and cytological studies describing differences between normal and malignant blood cells. There is no test tube reaction which would indicate malignancy. The differences noted between normal and malignant have been proven to be quantitative rather than qualitative and it is uncertain whether they are related to the cause of the malignant transformation or result from it. Claude fractionated extracts of malignant rat lymphocytes by means of centrifugation and has shown that ribose nucleic acid occurs in the cytoplasm of the leukemic cells only in association with formed elements of relatively large size, namely, microsomes and mitochondria (15a). These exist as physical entities and their structure is well shown by the electron microscope (16).

Tissue cultures. The abnormal characteristics of leukemic cells of mice are maintained in tissue cultures. Each cell type has a characteristic motility and this may help to identify the cell type (67). In addition, each malignant line of the same cell type studied proved to possess individual characteristics perpetuated in tissue cultures. The leukemic lymphoid cells studied (67) migrated from the explants more slowly than normal cells, showing characteristics of malignant cells in that they were larger and had larger and more granular nuclei than normal lymphocytes. The malignant monocytes studied (67) behaved in tissue cultures as the epithelioid cells of the tubercle. They were phagocytic, ingested red blood cells and formed bilirubin. Their nuclei were somewhat larger and had more chromatin granules than those of normal epithelioid cells (67).

Lewis and Lewis, and Bichel cultured malignant lymphocytes of mice over periods of months but no one has succeeded thus far in growing in vitro normal myeloblasts or lymphoid cells while leukemic cells seem to grow with ease.

Earlier investigators (cf. 33b) described maturation of human leukemic myeloid cells in vitro in the manner of normal immature myeloid cells. This observation, partially confirmed by Israëls is conflicting with those made on myeloid leukemia of mice and since it suggests a maturation defect as the basic

change in human leukemia, the discrepancy requires a clarification. Israëls goes as far as denying the applicability of experimental studies in rodents to human leukemias citing opinions and observations in favor of the non-neoplastic nature of human leukemias, but none of these appears conclusive. This investigator has cultured cells from the bone marrow from one case of myeloid leukemia and three cases of monocytic leukemias and noted maturation of the presumably leukemic cells. Such studies are liable to the pitfalls of interpretations made on the basis of fixed and stained preparations; they should be documented with sections and motion picture records.

The chorioallantoic membrane of the developing chick embryo provides a medium in which human leukemic leukocytes may grow for a short period (93a). Thiersch noted the development of hyperplastic osteosclerosis in fowls which hatched after inoculation with material from human leukemias. None of the fowls or embryos had leukemia. The development of these bony changes may not be a mere coincidence (cf. 56).

Metabolism of leukemic cells. Most earlier studies are subjects of a technical error. The circulating blood is seldom if ever a pure culture of malignant blood cells, and their number is variable and uncertain. Leukemia often produces a leukemoid change and malignant blood cells may be indistinguishable in appearance from their immature normal homologues. In working with tissues the non-malignant framework cells and blood components and other residual normal or injured cells of the host have not been given proper consideration (11, 12b, 111).

The anaerobic glycolysis of lymph nodes and spleens with induced leukemias showed no definite metabolic distinctions between control and leukemic mice irrespective of the type and extent of infiltration. The anaerobic glycolysis of livers, however, showed a two- to eight-fold increase depending upon the extent of infiltration. All three types of analysis, metabolic, histologic and transmission concurred, indicating a relatively sudden onset and rapid progress of leukemia a few months after painting the mice with methylcholanthrene (11).

The respiratory quotients of all tissues examined were below unity. The aerobic glycolysis values of preleukemic, leukemic and leukemoid lymph nodes, spleens and livers were 50 per cent to 100 per cent above normal. No evidence was obtained for the existence of a preleukemic metabolism different from normal. In the opinion of Burk the metabolism of induced mouse leukemia seems to represent the lowest limit of a type of metabolism of a malignant tissue (11).

The presence of a neoplasm influences a wide variety of biochemical processes in host tissue remote from the site of the tumor (12a, 12b, 111). The leukemic state may alter the metabolism of liver cells (48). Hall attributed the increase in anaerobic glycolysis rate of livers of leukemic mice to two factors: *a*, the presence in the liver of malignant lymphocytes that are carrying on anaerobic glycolysis at a rate similar to that of the same lymphoid cells in the lymph nodes; *b*, to a considerable increase in the glycolytic activity of the parenchymal cells of the liver. The latter appears to be due to the leukemic state and may be independent of the presence of leukemic cells in the organ.

Normal bone marrow cells other than erythrocytes fulfill all but one criterion of a tumor type of metabolism of Burk, the relatively high R.Q. (about 0.96) distinguishing them from malignant cells (112). The available data on R.Q. values are insufficient for generalization and the studies of Warren (112) indicate that no known metabolic reactions serve to distinguish malignant from non-malignant hemopoiesis.

Normal cells appear to utilize less pyruvate and usually convert a greater proportion of pyruvate to lactate than leukemic cells (1).

Histogenesis of the leukemic cell. The ease of induction of leukemia and the availability of strains in which the disease is almost inevitable at a given age, enable a study of its histogenesis. Unfortunately, observations on the interrelation of cell types on the basis of appearances in fixed and stained preparations proved undependable and so are these histological studies on the origin of leukemia.

The tissue culture studies and transmission experiments indicate that the leukemic cells are capable of proliferating indefinitely, and it is conceivable that the malignant change affects the homologous cell: the lymphocyte, myeloblast or monocyte. However the reticular framework cells of the blood-forming organs are widely regarded as the source of both normal and leukemic cells. Potter et al. (94) described a localized reticulum cell hyperplasia in the high leukemia strain C58, preceding leukemia. This is particularly conspicuous in the medullary regions of the lymph nodes and perivascular regions of the liver, and is absent in low leukemia strains examined at comparable ages. They concluded that the first malignant lymphocytes are produced by reticulum cells. Since this reticulum cell proliferation is widespread it is assumed that not all become sources of leukemic cells. In the low leukemia strains studied by the reviewer, reticulum hyperplasia is common with advancing age (35b) and is frequently associated with non-malignant extramedullary hemopoiesis. There seems to be an association between hemopoiesis and reticulum hyperplasia but it is possible that the same stimulus provokes both independently, or that the reticulum cell proliferation is secondary. One of the most common sites of reticulum cell hyperplasia is in the liver, yet this organ is an extremely rare site of lymphosarcoma and may show only scant involvement in the early stages of leukemia. In the Ak stock the thymectomy experiments suggest that the thymus is the common site of the primary leukemic change. The origin of malignant blood cells in reticulum requires better evidence than hitherto presented.

The best index for the leukemic character of hemopoiesis in a given organ is a positive bioassay. Bioassays made during the preleukemic phase of spontaneous or induced leukemia have disclosed the presence of malignant blood cells not suspected on histological examination. The solution of this problem may come from further systematic studies with non-leukemic and leukemic strains correlating histological changes with bioassays.

THE GROWTH OF LEUKEMIC CELLS. Neoplastic cells are usually designated as autonomous or unrestrained but these are only relative terms. Physiological

and other stimuli influencing normal hemopoiesis influence to a limited extent the growth of leukemic cells and numerous recent investigations are concerned with factors influencing their growth. Unfortunately, many of these investigations are of little value because of the lack of consideration of the growth interference occurring when there is a genetic differential between the leukemic cells and the host on which they are grafted.

Genetic and antigenic considerations. The growth of a leukemic cell in a host other than that in which it has arisen is subject to a restraint, varying in intensity with the magnitude of the differential between grafted cells and the hosts. The following are illustrative examples of the difference between the genetics of spontaneous and transmitted leukemia. In C3H mice, leukemia is almost non-existent, yet this strain is highly susceptible to the growth of cells arising in the high leukemia strain Ak (36). F1 hybrids between high and low leukemia strains have a moderate or low incidence of spontaneous leukemia, yet they are with rare exceptions as susceptible to leukemias of the high leukemia strain as members of the latter (36). Accordingly, experiments dealing with the genetics of transmissible leukemia have little, if any bearing upon problems of inherited susceptibility to this disease. Experiments on the induction of resistance to transplantable leukemia and modification of such resistance by chemical, physical or other agents are concerned with factors of transplantation and have little if any bearing on the course of spontaneous leukemia; e.g., a transmissible rat leukemia can be grafted on only about 80 per cent of rats of the stock of origin indicating the existence of a differential between host and leukemic cells (79b). Embryonic tissue produces resistance to grafts in about 85 per cent of the rats treated. Several chemical agents as well as x-rays are capable of breaking down this resistance (79b). Such experiments are not concerned with the growth of spontaneous neoplasms.

Leo Loeb, in a classical monograph on individuality has described the reaction of the body to alien cells introduced. The individual differential acts as inflammatory irritant and provokes the formation of antibodies in the new host. Tumors grafted on unrelated mice seldom take, and if they take they often regress. Tumors which take in unrelated mice often tend to regress after application of an agent somewhat damaging to the tumor cells. Such experiments raise the questions: Have the agents used altered the differential between host and recipient cells, depressing the reactivity of the host or increasing the antigenicity of the graft, or have they acted as a growth stimulant or depressor and could they act also in the absence of antigenic differentials against a spontaneous malignant tumor. Transplantation of leukemias with a differential between host and grafted cells contributes to an understanding of the individuality of the leukemic cells but does not directly disclose factors influencing the growth of spontaneous neoplasms. In some cases, immunization processes are apparent only after the growing neoplasm has been completely extirpated (Loeb, p. 405). Agents injuring the tumor may augment immune reactions resulting in a "cure" or decrease of the per cent of takes.

Accordingly, the most conclusive experiments on growth modifiers are those

conducted on spontaneous or induced neoplasms. Unfortunately, the leukemias of animals are not recognized early enough. The nearest to an ego is an identical twin, and the nearest to an identical twin are littermates of inbred mice. Continued inbreeding leads to homozygocity though the latter may never be attained. The use of such strains proved trustworthy in orientation studies for growth modifiers.

Unfortunately, leukemic cells may change in the course of successive transplantation. The cells often gain in proliferative vigor and in the course of years or months all chronic transmissible strains of the reviewer have ended up as acute, producing a fatal disease in 5 to 12 days instead of months. At the same time there was but slight or no change in the appearance of the leukemic cells.

In the course of transplantation by MacDowell et al. in inbred strains of mice, antigenic differences became evident between grafted cells and hosts with production of immunity against these cells. This immunity did not interfere with the genesis and growth of leukemic cells in "immunized" mice (81a).

MacDowell et al. have noted that by introducing small numbers of leukemic cells arising in the homologous strain the mouse may survive and tolerate increasingly larger numbers of cells. Hosts susceptible to tumor grafts can be protected also passively by treatment with minced tissue containing living cells from an actively immunized animal. Normal and embryonal tissue from an unrelated mouse may also confer immunity. Tissue of the same genetic constitution as the host is ineffective. Circulating antibodies play an important part in immunity to transplants of leukemic cells (44).

Leo Loeb believes that there is no individuality differential between normal and cancer cells of the same animal and immunotherapy of neoplasms is, therefore, not feasible. If there is a subtle genetic difference between a spontaneous neoplastic cell and the homologous normal cell of its host, as is suggested by proponents of the somatic mutation theory, it is not of an order to be antigenic in the host; otherwise it would be destroyed. It is possible that limitations of spread in spontaneous leukemias and failure to circulate in the blood stream in large numbers are due to slight antigenicity, but even this has thus far not been proved. The numerous immunity studies to prepare, by specific absorption, a solution acting specifically on malignant cells have been carried out with grafted tumors, and are, therefore, not applicable to spontaneous disease.

Blood cells would seem to be particularly suitable for the study of specific antibodies for malignant tissue proteins but such studies have thus far not been reported. Opinions on the subject are varied. MacDowell et al. consider immunization against spontaneous leukemic cells a possibility while for reasons outlined, the reviewer regards this as most unlikely.

In evaluating the experiments on modification of the growth of leukemic cells most informative are those conducted with spontaneous leukemia. Large numbers of spontaneous cases are necessary for the study of each factor because different lines of leukemic cells of the same type may vary widely in reactivity

to the same agent (31). This is well known for human neoplasms; e.g., the hormonal response of prostatic carcinoma seems to be lasting in less than 25 per cent of the cases, and is transient in 75 per cent. Next in value are studies made with transmissible leukemias arising and transplanted in a highly inbred strain. Leukemic cells vary greatly among themselves; hence, the use of several different lines is desirable. A given line of transmissible leukemia reacts uniformly to the same agent. For preliminary experiments, transmitted leukemias in well inbred lines may be more suitable than spontaneous leukemias. The onset and course of the leukemia is known with transmitted leukemias, while spontaneous leukemias are only available at the advanced stage and their course is unpredictable. The greater the differential between grafted cells and hosts, the less informative is the experiment. Tendency for regression or failure to take in 100 per cent of the animals (discounting technical errors) are indications of individuality differentials.

Hormonal growth factors. The female sex hormones and adrenal cortical hormones undoubtedly affect lymphopoiesis. There is ample evidence indicating that the latter inhibit the growth of some malignant lymphocytes while the former have not been tried on a comparable scale.

Adrenal cortical hormones. Heilman and Kendall have shown that 11-dehydro-17 hydroxycorticosterone (compound E) produces atrophy of thymus and of lymphoid tissues. When given to mice with a transmissible lymphosarcoma, it caused rapid regression of the growth, but in most animals regression was incomplete and temporary, and after an interval of a few days or weeks the growth recurred. Gradually the leukemic cells became refractory to this hormone and eventually the tumor caused the death of every mouse. Resistance to the hormone resided in the tumor cell for the resistant tumor transplanted to another animal remained resistant. A few mice which were given the cells of this tumor by intravenous inoculation were also treated with compound E but prolongation of life of the mice with systemic leukemia was slight. A reticulum cell sarcoma did not respond at all to this compound. Several transplantable mammary carcinomas were tested and only slight retardation in growth was observed.

In experiments on the effect of adrenalectomy on the susceptibility to transmitted leukemia, Murphy and Sturm used middle aged rats of a substrain of Wistar rats that are only partially susceptible to leukemia arising in this strain. Only 46 per cent of 32 inoculated animals developed the disease. Adrenalectomy has raised this figure to 90.3 per cent (79a). This effect may be entirely nonspecific and may mean no more than lowering the resistance of the host to grafts. Adrenalectomy produces stimulation of lymphoid activity and the latter should raise resistance to tumors, according to the theory of Murphy (78a). Sturm and Murphy (107) attribute the greater receptivity of adrenalectomized rats to transplanted leukemia to the same factor which stimulates lymphoid tissue. It is possible, however, that the effects of adrenalectomy are non-specific in that the growth of other neoplasms would also be enhanced by adrenalectomy. In other words, this drastic procedure might break down the

differentials existing between host and grafted cells. The resistance of adrenalectomized rats is known to be markedly depressed to toxins, poisons, bacterial and protozoan infections (92).

The inhibitory effect of adrenal cortical and pituitary adrenotropic hormones on the growth of transplanted leukemic cells was tested by Murphy and Sturm under the condition mentioned. From 0 to 10 per cent of inoculated controls failed to come down with leukemia as contrasted with about 20 to 60 per cent of those receiving the hormone (79c).

Female sex hormone. Experiments on the relation of sex hormones to the induction of leukemia already reviewed do not suggest that these hormones, given in physiological amounts have a marked effect on the proliferation of the malignant cells. There are no reports on the effect of subfatal quantities of these hormones on the growth of leukemic cells. Stilbestrol, given in large quantities, brings about atrophy of thymus followed by that of other lymphoid tissues and it retards the growth of mice (41). Estrogenic hormones are one of the few substances which will affect the lymphoid tissue in the absence of adrenals (101).

Little is known about the effect of other hormones on malignant blood cells. Induced myxedema did not alter the course of chronic lymphoid or myeloid leukemia in man (91).

Nutritional factors. Quantitative restriction of a qualitatively adequate diet resulted in a slight prolongation of life with transmitted lymphoid and myeloid leukemias, but death from leukemia was not prevented (31). Underfeeding was begun early in the course of the disease and was about as drastic as the mice could stand. Nevertheless, prolongation of life was less than that resulting from application of x-rays, benzol or arsenicals (31).

Nursing influence. The milk of certain lactating female mice promotes the growth of some transplantable lymphoid and myeloid leukemias in normally refractory mice (63). This influence was present throughout the period of lactation and was maintained through three generations in normally refractory mice. The factor is present in saline extracts from liver, spleen or mammary gland; dialyses through parchment membrane resists 85°C. for 20 minutes and 50 per cent glycerol for 30 days (63). The existence of such a factor analogous to the agent of mammary carcinoma is a challenge to the genetic theories of transplantation and throws new light on growth factors of neoplastic cells.

Chemical growth modifiers. (Experimental chemotherapy of leukemia.) Following the ideas of Peacock and Beck and of Haddow that a carcinogen may also inhibit the growth of neoplastic cells, Stamer and Engelbreth-Holm (105) tested the effect of intravenous injection of a colloidal suspension of 9,10-dimethyl-1,2 benzanthracene on transmissible leukemia. Small doses of this substance prolonged the life of the mice with transmitted leukemia and a large dose actually controlled it in all of five mice that had been given this treatment. It is noteworthy that this dose is fatal by the subcutaneous or intraperitoneal routes and that no tumors appeared after intravenous injection of this carcinogen. These experiments require confirmation and extension. Are other transmitted

and spontaneous leukemias affected by 9,10-dimethyl-1,2 benzanthracene; do other more potent leukemogens have a similar effect; are compounds related chemically but lacking carcinogenic potencies also effective in controlling the growth of leukemia; what is the action of this compound on normal hemopoiesis—are some of the questions to be answered by experiments.

Flory et al. (31) have tested numerous organic compounds using different strains of transmitted leukemia and found that different leukemias respond differently to the same agent but the effect of a given chemical on a given strain is fairly constant. Benzol doubled or tripled the life span of mice with transmissible chloroleukemia and occasionally arrested the disease. Its effects were manifest even when administered during the advanced stage of the disease. Other types of leukemias tested responded little if at all to benzol and no benzol derivative was found that would be more effective than benzol (31). Knowledge of the fate of benzol in the body and its mode of action might aid in the extension of these findings. Potassium arsenite prolonged somewhat the life of mice with certain transmissible leukemias. No organic arsenical tested had a stronger effect and none prevented death from leukemia (31).

Chambers et al. studied the effect of several compounds which inhibit the respiration of malignant lymphoid tissues in vitro, on the growth of normal and malignant lymphoid cells in tissue culture. These include N,N,N^1,N^1-Tetramethyl-o-Phenylenediamine which was about seven times more toxic to malignant than to normal lymphoid cells. In general, the malignant cells were of the more immature type and it is possible that immaturity per se was the cause of excessive sensitivity (14). In tissue cultures lymphoid sarcoma cells of rats were more susceptible to heptanal sodium bisulfite methylsalicylate than normal lymphoid cells (Cameron et al.). It remains to be shown whether these short-term tissue culture studies yield results that parallel those in vivo. Heptylaldehyde bisulfate strongly depresses respiration of lymphocytes in vitro but it had no effect on lymphomatous tumors of rats and on leukemic patients (102). Colchicine temporarily inhibited the growth of transmitted lymphosarcoma (68, 117).

Physical agents. Lymphomatous tissue retains more radioactive phosphorus than normal lymphoid tissue from 1 to 19 days after administration of tagged phosphorus. The exchange (specific activity) of phosphorus in the lymphoma studied was twice as great as in normal lymphoid tissue (100). In rapidly growing lymphomatous tissue the nucleoprotein fraction was found to contain 90 to 95 per cent of the total nuclear radioactive phosphorus. From the rate of its uptake it was calculated that a new lymphoma nucleus was synthesized once every 27 hours. The great accumulation of P32 in malignant cells is attributed to an increase in mitotic activity (71a). Warren found that administrated radioactive phosphorus is retained in greatest amounts in those organs of leukemic patients which are heavily infiltrated with leukemic cells. Slowly metabolizing tissues contain but little phosphorus. The distribution of radioactive phosphorus in man is of approximately the same order as in rodents (113a). Radioactive phosphorus is a useful means of therapy of leukemia

because of its preferential concentration in rapidly proliferating cells, hence in leukemic infiltrations (113b). If given many days before death, the concentration of P32 is greatest in osseous tissues while if given just before death, most is found in the bone marrow, lymph nodes, spleen and liver. This suggests that it is first utilized by more rapidly metabolizing tissues and by those most frequently infiltrated with leukemic cells and later by bones (28). Numerous other studies with different radioactive elements, neutrons and x-rays have been reviewed by Lawrence.

X-rays. Of all agents tested by Flory et al., x-rays had the broadest range in destroying malignant blood cells. They prolonged the life of most leukemic animals but in no case has it been possible to destroy all malignant cells without doing fatal damage to normal organs.

Following exposure to x-rays and to neutrons there is a sharp increase in abnormal mitoses 3 hours after exposure and a greater increase at 12 hours. As the dose of the x-rays increases the percentage of mitoses decreases indicating (Marschak, 71b) that the damage results from a single encounter of the x-rays with a sensitive part of the chromosomes. Unlike those of other cells studied, chromosomes of a mouse lymphoma are sensitive to x-rays and neutrons not only at the onset of the prophase but also during the period of the resting stage where the sensitivity is even greater (71b). Since neutrons are more efficient than x-rays in producing chromosome abnormalities in the nuclear resting stage, Marshak predicts that they will produce regression of neoplasms resistent to x-rays (71b).

Cold. Exposure of starved leukemic mice to cold at 6–8°C., during 5 to more than 24 hours, caused degeneration and disappearance of leukemic lymphocytes but the leukemic infiltrations reappeared in all animals that survived this drastic treatment (26). It is noteworthy in this connection that malignant lymphocytes resist freezing (9).

AVIAN LEUKOSIS AND THE VIRUS PROBLEM. The virus theory of leukemia is based mainly on experiments with avian leukemia (leukosis) and the latter will, therefore, be briefly reviewed. The problems of avian leukosis are identical with those of avian sarcoma. Numerous viruses have been isolated which produce either leukosis or sarcoma or both. The range of each virus is constant, however, under usual conditions. Earlier experiments suggesting a change in their affinities have not been confirmed. Duran-Reynals has shown that the agent of chicken sarcoma is capable of profound modification when passed through another species but it is uncertain how far leukemogenic properties can be acquired or lost in this manner.

Three main types of avian leukosis are known; erythro- and myeloleukosis, which are usually caused by the same agent, and lymphoid leukosis, which can be transmitted with difficulty or not at all. The latter occurs in many forms and is probably caused by several different viruses. An avian lymphoma readily transmissible by cellular grafts, has been described by Olson. No cell free agent has been demonstrated in this lymphoma but no intense search seems to have been made for such an agent. Numerous investigators have induced sarcoma in chicks with hydrocarbons but it is debatable if these tumors

contain an agent (virus). Nobody has described the production of leukosis in chickens by pure carcinogenic chemicals. Immunological studies suggest the presence of a virus in a chemically induced chicken tumor that is not transmissible by cell free material; sera of fowls bearing such tumors neutralize a leukosis-sarcoma agent (Gottschalk). The direct demonstration of an agent in chicken leukosis or sarcoma is sometimes difficult and the studies of Rous et al. and of others indicate that no avian tumor or leukosis can be certified to be free from such agents. A non-neoplastic change, osteosclerosis, is believed to be caused by some agents of leukosis (56).

A characteristic form of avian leukosis associated with infiltrations in the nervous system commonly called fowl paralysis or neurolymphomatosis is of great economic importance and has been studied extensively in numerous places. It appears to be sometimes neoplastic and sometimes inflammatory in character with predominance of small lymphocytes. Numerous transmission experiments yielded suggestive evidence of its transmissibility until recently Blakemore discovered that it is caused by a filter-passing agent which produces a non-fatal myocarditis in baby chicks. Neurolymphomatosis is conceived to be a late phase of the same disease. There are distinct hereditary differences in susceptibility to agents of fowl leukosis and sarcoma (54, 114). Leukemic cells placed on the chorioallantois produce leukosis in chick embryos but the cell free virus does not although it persists in the embryo (93b). In tissue cultures the agent of fowl leukosis could be grown in the presence of fibroblasts (21). Thus this agent is capable of multiplication in the presence of cells which it does not render neoplastic.

For years, the discussion centered around the question whether the cause of avian leukosis was a virus similar to that of other contagious diseases or an autocatalytic endogenous selfperpetuating substance inducing other cells to an irreversible neoplastic change. Neither the morphological characteristics nor behavior in the body would distinguish a virus-induced leukosis or other neoplasm from one seemingly free from virus. There are strains of avian leukosis and sarcoma readily transmissible by cell free agents, strains transmissible by cells only and all gradations in between. This virus can be conceived as part of the reproductive matter of the abnormal cell (20, 19) and the fact that it is separable from the cell makes it a useful tool for the study of cell determinants. Normal cells also contain in their cytoplasm sedimentable substances of the order of magnitude of viruses. Chemical and immunological studies suggest a similarity between virus and normal heavy tissue particles (15b). They are carriers of numerous enzymatic activities (cf. Kabat). It is possible that the virus is a modified normal heavy cytoplasmic substance (Claude) or that the so-called virus thus far analysed is essentially normal heavy material and only a small fraction of it is virus (Kabat). The trend is to consider these viruses as essential but altered components of the heavy material present in normal cells and to regard them as autosynthetic as are genes and some enzymes. The carcinoma protein is a modified normal protein with changed enzymatic activity occasionally containing a "reproductive particle" (V. R. Potter).

The experiments of Duran-Reynals on modification of an avian neoplastic

virus through passage of different species are explained by assuming that the new environment provides different "building blocks," causing a reconstruction of the virus, the altered virus becoming dominant in the new host (Dixon). Needham compares carcinogenesis by virus with neural plate induction. The latter could also be propagated by cell free extracts. The synthesis of further supplies of evocators may be stimulated autocatalytically like trypsin from trypsinogen or plant viruses in their hosts.

In the light of these investigations, neoplasia in mammals can be explained by a similar cytoplasmic disturbance yielding a virus-like agent. Engelbreth-Holm and Frederikson described experiments suggesting such an agent in mouse leukemia, but their findings still await confirmation. This agent, they thought, is rapidly inactivated by oxidation.

THE SOMATIC MUTATION THEORY. The idea that the neoplastic change is analogous to a somatic mutation has had its proponents and opponents for several decades (36). All somatic cells of an organism are believed to have an identical chromosomal make-up with a complete set of genes responsible for the organismal differentials. Differentiation or transformation of one cell type into another is not accompanied by chromosomal changes. When the somatic mutation theory was proposed, only chromosomes were regarded to be hereditary components of the cell. Subsequently, cytoplasmic materials have been recognized that are self-reproducing and important in determining the characteristics of the cell. The term somatic mutation will be used here in a broad sense, meaning changes in any component of the cells which is self-reproducing and determines cellular characteristics. An acceptance of the idea that the neoplastic change is analogous to a somatic mutation requires proof that the change resides in the chromosomes or in reproductive cytoplasmic particles. No such changes have thus far been noted in the chromosomes of leukemic cells, and nothing is known about the hypothetical reproductive particles in the cytoplasm of these cells.

Indirect evidence for a genetic basis of the leukemic alteration is suggested by transplantation tests. Transplantation tests in highly inbred strains are believed to yield information on the genetic character of the cells. Hybrids of known genetic composition have a greater assortment of chromosomes than homozygous inbred mice and transplantation tests in F1 hybrids yielded data of greater value. It has been known that grafts of normal cells of F1 hybrids succeed 100 per cent in F1 hybrids but not at all in parental mice. This was verified for spleens of the F1 hybrids of a high (Ak) and a low (Rf) leukemia strain in which these transplantation experiments were made (58). Numerous lines of leukemic cells arising spontaneously in F1 and other hybrids of these strains were grafted on parental and on hybrid mice. Unlike normal spleens they proved readily transplantable to mice of the parental high leukemia mice (99). Most leukemias induced by methylcholanthrene in F1 hybrids of these strains could not be grafted on parental mice but some could be grafted on either or on both parental strains (34).

These results are explained by Loeb by assuming that F1 cells contain, in addition to the gene sets derived from the parents which determine the "or-

ganismal differentials" of the leukemic cells, an intrinsic stimulus (named by him Gi—intrinsic growth factor) which is derived from the parent susceptible to the development of spontaneous leukemia. This intrinsic stimulus is responsible for the conversion of normal lymphoid cells into leukemic cells. In the high leukemia strain the existence of an auxiliary extrinsic growth promoting factor (Ge) is postulated, which makes it possible for leukemic cells to multiply. This hypothetical Ge factor is not identical with the nursing influence of Law (already described) since the former is transmitted by both mother and father. It may represent gene hormones through which in the course of embryonal development, gene effects may be transmitted to the recipient tissues (Loeb, p. 383).

Not all leukemic cells arising in F1 hybrids behaved in the manner described. Kirschbaum and Strong (62b) noted that leukemic cells arising in CBA x F hybrids behaved in transplantation tests as normal spleens.

Mutations are believed to be sudden while the neoplastic change is stated to be a step by step non-genetical process. This, however, may be more apparent than real, as was recently pointed out by Mottram for cancer cells, the more rapidly proliferating cell of a heterogeneous malignant cell population gradually outnumbering the less rapidly growing cell lines. Another objection to the somatic mutation hypothesis is that malignant cells frequently change in the course of transplantation. Malignant cells, as other living cells, are subject to inheritable modifications but, since the use of inbred strains of mice, the only consistent change noted in lines of malignant cells is that in growth vigor but not a change in character.

When a normal cell turns malignant it can survive and proliferate only when the deviation in its components is not of sufficient magnitude to create a new antigen. The formation of a new substance is not necessary for antigenicity; structural differences create antigenic differences. Thus, the abnormal in the neoplastic cell need not be a plus material (virus or agent) but a changed nucleoprotein complex. This is in essence the theory of numerous investigators. Malignancy in relation to organization and differentiation has been reviewed recently in this journal (5).

The transplantation experiments described demonstrate differences between normal F1 cells and neoplastic F1 cells. In order to explain this difference, more knowledge is required of chromosomes of blood cells and of the assumed cytoplasmic reproductive cell determinants. For such studies uniform populations of normal and neoplastic blood cells are now available, and the recent advances in histochemistry, electron microscopy, the introduction of tagged elements, etc. might enable a more precise definition of the abnormal in the neoplastic cell than has hitherto been possible.

CLOSING COMMENTS

Contrary to general belief, control of a disease may either precede or follow knowledge of its cause and character. Basic unknowns can be investigated irrespective of their logical sequence: factors regulating normal hemopoiesis, the causation of leukemia, the fundamental change in leukemogenesis, the

character of the leukemic cell and the factors influencing its growth. Each of these major problems can be resolved into a mosaic of minor questions. Some of these and the facts gathered to answer them during the past few years have been reviewed. Solidity of facts gives strength to a theory. Major discoveries have been announced on the basis of a single or few small experiments and left abandoned. Leukemia and leukemoid tissue reactions have often not been distinguished. Factors modifying the growth of grafts have been confused with those of the origin and growth of leukemic cells in their native host. Heredity, age, sex, nutrition and maternal influences are some of the ever present and often forgotten variables modifying, if not determining the action of an agent chosen for study. In interpreting and analyzing experimental results, there is often a lack of integration with data of others and the generalization made is sometimes not warranted.

The essence of the leukemic change appears as follows: diverse extrinsic and genetic factors acting on normal hemopoietic cells result in the formation of an abnormal immature blood cell with a heightened growth force. This, the leukemic cell, possesses individual characteristics which distinguish it from immature normal cells of its host and from other leukemic cells. Its most important characteristics are: failure of complete differentiation and ability to overcome the forces which restrain the growth of normal immature cells of its kind, so that it invades tissues and organs and ultimately kills its host. It may remain at the site of origin or migrate around the body multiplying at favored sites. The basic change is intrinsic in the cell since it is maintained in tissue cultures. The change involves the enzymatic and metabolic capacities of the cell but no such single qualitative change is known which would identify a neoplastic blood cell. The basic change may be due to *a*, an alteration of the reproductive material of the cells, either cytoplasmic or nuclear (the theory of somatic mutation), or *b*, to the presence of a selfperpetuating exogenous agent (virus theory in a restricted sense), or *c*, to an endogenous growth factor multiplying autocatalytically, causing abnormal differentiation.

The origin of the somatic mutation theory is obscure. It is associated with the name of Boveri but had many proponents. The virus theory is supported by the work of Rous, Andrewes, Bittner etc. (cf. 97) and that of abnormal differentiation related to organizer effects is suggested by Loeb, Murphy, Needham, etc. (cf. 85). The current trend is for the integration of these theories as it seems that the frontiers existing between studies of heredity, development and infection are only technical and arbitrary (Darlington).

I gratefully acknowledge the extended support given to investigations in this Department by The Lady Tata Memorial Trust, The International Cancer Research Foundation, The Jane Coffin Childs Memorial Fund for Medical Research, The Anna Fuller Fund; and by The Anonymous Donor who initiated these studies.

REFERENCES

(1) Abels, J. C., F. L. Jones, L. F. Craver and C. P. Rhoads. Cancer Research **4:** 149, 1944.
(2) Barnes, W. A. and R. K. Cole. Cancer Research **1:** 99, 1941.
(3a) Barnes, W. A. and J. Furth. Am. J. Cancer **30:** 75, 1937.
(3b) Barnes, W. A. and J. Furth. Am. J. Path. **16:** 457, 1940.

(4) BARNES, W. A. AND I. E. SISMAN. Am. J. Cancer **37**: 1, 1939.
(5) BERRILL, N. J. Physiol. Rev. **23**: 101, 1943.
(6a) BICHEL, J. Ztschr. f. Krebsforsch. **48**: 92, 1938.
(6b) BICHEL, J. Monograph of Univ. Press of Aarhus, Denmark, 1939.
(7a) BIELSCHOWSKY, F. Brit. J. Exper. Biol. **25**: 1, 1944.
(7b) BIELSCHOWSKY, F. Ibid. **25**: 90, 1944.
(8a) BLAKEMORE, F. J. Comp. Path. and Therap. **52**: 144, 1939.
(8b) BLAKEMORE, F. Proc. Roy. Soc. Med. **34**: 109, 1941.
(9) BREEDIS, C. J. Exper. Med. **76**: 221, 1942.
(10) BRUES, A. M. AND B. B. MARBLE. Am. J. Cancer **37**: 45, 1939.
(11) BURK, D., H. SPRINCE, J. M. SPANGLER, M. C. BOON AND J. FURTH. J. Nat. Cancer Inst. **3**: 249, 1942.
(12a) BURK, D. AND R. J. WINZLER. In Vitamins and hormones *II*, 1944.
(12b) BURK, D. AND R. J. WINZLER. Ann. Rev. Biochem. **13**: 487, 1944.
(13) CAMERON, G., C. J. KENSLER AND R. CHAMBERS. Cancer Research **4**: 495, 1944.
(14) CHAMBERS, R., G. CAMERON AND M. J. KOPAC. Cancer Research **3**: 293, 1943.
(15a) CLAUDE, A. J. Exper. Med. **80**: 19, 1944.
(15b) CLAUDE, A. Cold Spring Harbor Symposia on Quantitative Biology. Cold Spring Harbor Long Island Biological Association **9**: 263, 1941.
(16) CLAUDE, A. AND F. FULLAM. J. Exper. Med. **81**: 51, 1945.
(17) CLAUDE, A. AND J. B. MURPHY. Physiol. Rev. **13**: 246, 1933.
(18) COLE, R. K. AND J. FURTH. Cancer Research **1**: 957, 1941.
(19) DARLINGTON, C. D. Nature **154**: 164, 1944.
(20) DIXON, T. F. Nature **155**: 596, 1945.
(21) DOLJANSKI, L. AND M. PIKOVSKI. Cancer Research **2**: 626, 1942.
(22a) DOUGHERTY, T. F. AND A. WHITE. Proc. Soc. Exper. Biol. and Med. **53**: 132, 1943.
(22b) DOUGHERTY, T. F. AND A. WHITE. Science **98**: 367, 1943.
(23) DURAN-REYNALS, F. Cancer Research **3**: 569, 1943.
(24) ELLERMANN, V. The leucosis of fowls and leucemia problems. Gyldendal, London, 1921.
(25) ENGELBRETH-HOLM, J. Leukemia in animals. Oliver and Boyd, Edinburgh, 1942.
(26) ENGELBRETH-HOLM, J. AND H. ELTORM. Acta path. et microbiol. Scand. **20**: 346, 1943.
(27) ENGELBRETH-HOLM, J. AND O. FREDERIKSON. Acta path. et microbiol. Scand. Suppl. **37**: 145, 1938.
(28) ERF, L. A. Am. J. Med. Sc. **203**: 529, 1942.
(29) FARRIS, E. J. AND E. H. YEAKEL. Am. J. Path. **20**: 773, 1944.
(30) FISCHER, A. Virchow's Arch. **310**: 395, 1943.
(31) FLORY, C. M., J. FURTH, J. A. SAXTON, JR. AND L. REINER. Cancer Research **3**: 729, 1943.
(32) FORKNER, C. E. Leukemia and allied disorders. MacMillan, New York, 1938.
(33a) FURTH, J. J. Exper. Med. **69**: 13, 1939.
(33b) FURTH, J. Wisconsin University Press **106**: 125, 1939.
(33c) FURTH, J. AAAS Cancer Symposium, 1945. In print.
(33d) FURTH, J. Unpublished data.
(33e) FURTH, J. J. Gerontology. To be published.
(34) FURTH, J. AND W. A. BARNES. Cancer Research **1**: 17, 1941.
(35a) FURTH, J. AND M. C. BOON. Science **98**: 138, 1943.
(35b) FURTH, J. AND M. C. BOON. AAAS Cancer Symposium, 1945, in print.
(36) FURTH, J., M. C. BOON AND N. KALISS. Cancer Research **4**: 1, 1944.
(37) FURTH, J., R. K. COLE AND M. C. BOON. Cancer Research **2**: 280, 1942.
(38) FURTH, J., H. W. FERRIS AND P. REZNIKOFF. J. A. M. A. **105**: 1824, 1935.
(39) FURTH, J. AND N. KRUMDIECK. Am. J. Path. **16**: 449, 1940.
(40) GATES, R. R. Human genetics. In press.
(41) GARDNER, W. U., T. F. DOUGHERTY AND W. L. WILLIAMS. Cancer Research **4**: 73, 1944.

(42) Gardner, W. U., A. Kirschbaum and L. C. Strong. Arch. Path. **29**: 1, 1940.
(43) Gilmour, M. D. J. Path. and Bact. **45**: 179, 1937.
(44) Gorer, P. A. J. Path. and Bact. **54**: 51, 1942.
(45) Gottschalk, R. G. Cancer Research **3**: 649, 1943.
(46) Grand, C. G. Proc. Soc. Exper. Biol. and Med. **56**: 229, 1944.
(47) Haddow, A. Nature **154**: 194, 1944.
(48) Hall, V. E. Cancer Research **4**: 785, 1944.
(49) Heilman, F. R. and E. C. Kendall. Endocrinology **34**: 416, 1944.
(50) Heinle, R. W., J. T. Wearn, D. R. Weir and F. A. Rose. Ann. Int. Med. **17**: 902, 1942.
(51a) Henshaw, P. S. J. Nat. Cancer Inst. **4**: 513, 1944.
(51b) Henshaw, P. S. Radiology **43**: 279, 1944.
(52) Henshaw, P. S. and J. W. Hawkins. J. Nat. Cancer Inst. **4**: 339, 1944.
(53) Hirschmann, R. R., R. W. Heinle and J. T. Wearn. Proc. Soc. Exper. Biol. and Med. **58**: 5, 1945.
(54) Hutt, F. B., R. K. Cole, M. Ball, J. H. Buckner and R. F. Ball. Poultry Sc. **23**: 396, 1944.
(55) Israëls, M. C. G. J. Path. and Bact. **51**: 235, 1940.
(56) Jungherr, E. and W. Landauer. Conn. (Storrs) Agric. Exper. Sta. Bull. **222**: 5, 1938.
(57) Kabat, E. A. J. Immunol. **47**: 513, 1943.
(58) Kaliss, N. and T. Robertson. Genetics **28**: 78, 1943.
(59) Kaplan, H. S. and A. Kirschbaum. Proc. Soc. Exper. Biol. and Med. **55**: 262, 1944.
(60a) Kirschbaum, A. Proc. Soc. Exper. Biol. and Med. **55**: 147, 1944.
(60b) Kirschbaum, A. Yale J. Biol. and Med. **17**: 163, 1944.
(61) Kirschbaum, A. and H. S. Kaplan. Science **100**: 360, 1944.
(62a) Kirschbaum, A. and L. C. Strong. Am. J. Cancer **37**: 400, 1939.
(62b) Kirschbaum, A. and L. C. Strong. Cancer Research **1**: 785, 1941.
(62c) Kirschbaum, A. and L. C. Strong. Proc. Soc. Exper. Biol. and Med. **51**: 404, 1942.
(62d) Kirschbaum, A. and L. C. Strong. Cancer Research **2**: 841, 1942.
(63) Law, L. W. Cancer Research **4**: 257, 1944.
(64) Lawrence, J. H. Am. J. Roentgenol. **48**: 283, 1942.
(65) Lewis, M. R. Am. J. Cancer **29**: 510, 1937.
(66) Lewis, M. R. and G. B. Mider. J. Nat. Cancer Inst. **2**: 115, 1941.
(67) Lewis, W. H. and M. R. Lewis. In cooperation in Research. Carnegie Institute of Washington Publication **501**: 369, 1938.
(68) Lits, F., A. Kirschbaum and L. C. Strong. Am. J. Cancer **34**: 196, 1938.
(69) Loeb, L. Biological basis of individuality. Thomas, Springfield, 1944.
(70) March, H. C. Radiology **43**: 275, 1944.
(71a) Marshak, A. J. Gen. Physiol. **25**: 275, 1941.
(71b) Marshak, A. Radiology **39**: 621, 1942.
(72) Mayneard, W. V. and L. D. Parsons. J. Path. and Bact. **45**: 35, 1937.
(73a) Mercier, L. Compt. rend. Soc. de biol. **124**: 403, 1937.
(73b) Mercier, L. Ibid. **133**: 29, 1940.
(74) Miller, F. R., W. A. Hause and H. W. Jones. Proc. Soc. Exper. Biol. and Med. **50**: 115, 1942.
(75) Miller, F. R. and D. L. Turner. Am. J. Med. Sc. **206**: 146, 1943.
(76) Morton, J. J. and G. B. Mider. Science **87**: 327, 1938.
(77) Mottram, J. C. Am. J. Cancer **30**: 746, 1937.
(78a) Murphy, J. B. Monograph of Rockefeller Inst. for Med. Res. no. 21, 1926.
(78b) Murphy, J. B. Cancer Research **4**: 622, 1944.
(79a) Murphy, J. B. and E. Sturm. Science **98**: 568, 1943.
(79b) Murphy, J. B. and E. Sturm. Cancer Research **3**: 173, 1943.
(79c) Murphy, J. B. and E. Sturm. Science **99**: 303, 1944.
(80a) MacDowell, E. C. J. Heredity **28**: 131, 1937.

(80b) MacDowell, E. C. In Cooperation in Research. Carnegie Institute of Washington Publication **501**: 353, 1938.
(81a) MacDowell, E. C., J. S. Potter and M. J. Taylor. Proc. Nat. Acad. Sci. **25**: 416, 1939.
(81b) MacDowell, E. C., J. S. Potter and M. J. Taylor. Cancer Research **5**: 65, 1945.
(82) MacDowell, E. C., J. S. Potter, M. J. Taylor and E. N. Ward. Rec. Genetics Soc. America **9**: 160, 1940.
(83) MacDowell, E. C., J. S. Potter, J. Victor, M. J. Taylor, T. Laanes, E. N. Ward and M. P. Wintersteiner. Carnegie Institute of Washington Year Book **36**: 51, 1937.
(84a) McEndy, D. P., M. C. Boon and J. Furth. J. Nat. Cancer Inst. **3**: 227, 1942.
(84b) McEndy, D. P., M. C. Boon and J. Furth. Cancer Research **4**: 377, 1944.
(85) Needham, J. Biochemistry and morphogenesis. Cambridge Univ. Press, Cambridge, 1942.
(86) Oberling, C. The riddle of cancer. Yale Univ. Press, New Haven, 1944.
(87) Olson, C. Cancer Research **4**: 707, 1944.
(88) Opie, E. L. Proc. Inst. Med. Chicago **14**: 382, 1943.
(89) Parks, A. S. Physiol. Rev. **25**: 203, 1945.
(90) Parsons, L. D. J. Path. and Bact. **54**: 321, 1942.
(91) Paul, J. T., L. R. Limarzi and L. Seed. Am. J. Med. Sc. **206**: 625, 1943.
(92) Perla, D. and J. M. Gottesman. Proc. Soc. Exper. Biol. and Med. **28**: 475, 1931.
(93a) Pierce, M. Arch. Path. **34**: 538, 1942.
(93b) Pierce, M. Am. J. Path. **18**: 1127, 1942.
(94) Potter, J. S., J. Victor and E. N. Ward. Am. J. Path. **19**: 239, 1943.
(95) Potter, V. R. Science **101**: 609, 1945.
(96) Richter, M. N. and E. C. MacDowell. Physiol. Rev. **15**: 509, 1935.
(97) Rous, P. J. A. M. A. **122**: 573, 1943.
(98) Saxton, J. A., M. C. Boon and J. Furth. Cancer Research **4**: 401, 1944.
(99) Schweitzer, M. D. and J. Furth. Am. J. Cancer **37**: 224, 1939.
(100) Scott, K. G. Cancer Research **5**: 365, 1945.
(101) Selye, H. Brit. J. Exper. Path. **17**: 234, 1936.
(102) Singher, H. O., J. C. Abels, L. F. Craver and C. P. Rhoads. Cancer Research **4**: 444, 1944.
(103) Slye, M. Am. J. Cancer **15**: 1361, 1931.
(104) Stamer, S. Acta path. et microbiol. Scand. Suppl. **47**: 1, 1943.
(105) Stamer, S. and J. Engelbreth-Holm. Acta path. et microbiol. Scand. **20**: 360, 1943.
(106) Stoerk, H. C. and T. F. Zucker. Proc. Soc. Exper. Biol. and Med. **56**: 151, 1944.
(107) Sturm, E. and J. B. Murphy. Cancer Research **4**: 384, 1944.
(108a) Tannenbaum, A. Arch. Path. **30**: 509, 1940.
(108b) Tannenbaum, A. Cancer Research **2**: 460, 1942.
(108c) Tannenbaum, A. Cancer Research **4**: 678, 1944.
(109) Thiersch, J. B. Australian J. Exper. Biol. and Med. Sc. **22**: 57, 1944.
(110) Turner, D. L. and F. R. Miller. J. Biol. Chem. **147**: 573, 1943.
(111) Victor, J. and J. S. Potter. Brit. J. Exper. Path. **19**: 227, 1938.
(112) Warren, C. O. Cancer Research **3**: 621, 1943.
(113a) Warren, S. Cancer Research **3**: 334, 1943.
(113b) Warren, S. Ibid. **3**: 872, 1943.
(114) Waters, N. F. Poultry Sc. **24**: 259, 1945.
(115) White, F. R., J. White, G. B. Mider, M. G. Kelly and W. E. Heston. J. Nat. Cancer Inst. **5**: 43, 1944.
(116) White, J., G. B. Mider and W. E. Heston. J. Nat. Cancer Inst. **4**: 409, 1944.
(117) Williams, W. L. Anat. Rec. suppl. 1944.
(118) Wilson, R. H., F. DeEds and A. J. Cox, Jr. Cancer Research **1**: 595, 1941.

NEURAL MECHANISMS OF CUTANEOUS SENSE

GEO. H. BISHOP[1]

Laboratory of Neurophysiology, Oscar Johnson Institute, Washington University Medical School, St. Louis

Following numerous recent reviews of the cutaneous senses, the present effort may appropriately emphasize a particular aspect of sensory function. The underlying concept will be that all sensory messages, on their passage through nerves from the periphery to the center, must consist of patterns of non-specific nerve impulses. The consequent problem is then to correlate the pattern of the impulses with the sensation induced, considering the nerve impulses as the coin in terms of which we may evaluate sensation on the one hand, and stimulation on the other. These patterns can vary only in limited respects to account for variations in sensation.

First, different fibers, for instance large and small, may carry impulses of different energy value, i.e., of different denominations as coins. The significance of this is little understood. Second, any one fiber may fire at different frequencies, and for different durations, resulting in an accumulation or increase of sensory value in terms of temporal summation. Third, fibers ending at different points of the skin or even at the same point may be activated together, at the same or at different frequencies, resulting in spatial summation of sensory values, sometimes, as might be expected, indistinguishable in sense from temporal summation. No other variations would seem to be possible in the nerve-impulse patterns of sensory messages.

Two other types of specialization of function however exist. In the periphery the endings of the fibers are differently located, are differentially sensitive to different forms of energy, and have different properties appearing as adaptation, threshold, refractoriness, after-discharge, etc. The character of the pattern of impulses set up by any external stimulus is largely determined by these factors. Then centrally the different fibers impinge on different cells or areas of cortex, or make differential connections in some other characteristic manner, such that impulses however originated arriving over different fibers set up different sensory effects. It is further probable that at the synapses along the fiber pathways to the cortex, differences exist characteristic of different groups of fibers or of sensory pathways, modifying the pattern from the periphery which is received at the cortex. In addition to these specific differences in connections, the sensory effect is further dependent on the state of excitability or of activity of the center, usually conceived of in terms such as mental or emotional attitude, state of attention, memory, etc. It is also dependent on the state of the periphery determining the effects of a given stimulus on the pattern, for which terms of a corresponding elegance have not been elaborated. At both center and periphery the state of excitability and the consequent magnitude and form of response may be further altered by current events, that is, by the stimulus itself at the periphery, and by

[1] Assisted by a grant for research in neurophysiology from the Rockefeller Foundation.

the impulses arriving at the center. The final and overall effect of all these factors is the sensory process.

It will thus be obvious that skin sensation is more than skin-deep. In fact the only portion pertaining directly to the skin of the extensive nervous machinery which performs a sensory act is the peripheral skin "sense" organ. Even it determines only in a preliminary way what sensation is aroused by an external stimulus. The term "cutaneous sensation" is appropriate only to suggest that a class of discriminatory or affective states of the central brain is potentially but not necessarily aroused by certain agents acting on the skin. To these agents the skin is exposed by the incidental circumstance that it finds itself lying across the pathway of some physical force aimed at the organism in general. Further, sensation is only one of the ways in which the organism reacts, for it also reacts to the same stimulus at various levels of the nervous system beside the sensory.

Sensory pathways are in fact only a particular selection of the afferents to the central nervous system. They consist actually only of those afferents the results of whose excitation may reach consciousness. Sensation being registered in the brain, wherever along the pathway activation occurs, it is obviously more than the functioning even of these sensory afferents. It is therefore appropriate to examine first the criteria by which "cutaneous sensation" may be differentiated out of the complex of nervous system behavior.

The concept of modality. The four modalities of cutaneous sense usually recognized, touch, pain, cool and warm, are differentiated on a dual basis, subjective and objective, and this has given rise to some confusion in the defining of sensory categories. Moreover the flexibility of nervous system behavior leads to such variation in results of stimulation as to require rather arbitrary and perhaps artificial conditions of experimentation in the derivation of a formal classification. Centrally, the ability exists to accurately differentiate subjectively, *at least* four sensations in simple experimental situations; and peripherally, at least four corresponding classes of endings in the skin can be inferred, if not ultimately demonstrated histologically, reacting typically but not exclusively to different and characteristic forms of energy. It is a degree of correspondence between these subjective, peripheral and physical phenomena which gives meaning to the idea of modality. All other sensations aroused by cutaneous stimuli can then be defined as summations of these four components, in various combinations and intensities. Two difficulties arise in this connection.

First, more than four types of endings are present, as judged by punctate stimulation of the skin at least; and second, the various stimulation of one point, and presumably of one type of ending, may arouse different sensations. It would seem logically necessary for instance to include in the pain modality the rather pleasant sensation derived from scratching an itch, which must stimulate pain endings, and probably to assign itch itself to the result of a particular pattern of activation of pain endings. Again the vibratory sense, although distinguishable from either touch or pressure, probably employs both pathways in a characteristic pattern of activation, and several distinguishable varieties of touch following point stimulation of skin are included in one modality merely because they seem

to have, subjectively, a common quality. Still other sensations are more obviously the result of a more complex pattern of multiple stimulation; wetness, smoothness, etc. The selection of four fundamental modalities for the classification of cutaneous sensations is thus a compromise between simplicity and confusion, and between specificity of stimulation and flexibility of central interpretation.

One may then define pragmatically the modality of warmth as that sensory process set off with least energy when heat only is applied as a stimulus to the normal or average subject, over a considerable skin surface. Cool can be defined as the sensation aroused from the stimulation by an adequate stimulus of certain discrete points usually occurring in groups deep in the skin, and which can be identified as those responding preferentially when lowered temperature is the only stimulus. Pain as a cutaneous modality is the response to any stimulus at certain discrete points in the skin, characterized normally by an avoidance or withdrawal reaction if the stimulus is sufficiently intense, and arousing a central sensation which is its own definition. Below this intensity of stimulation these sense organs arouse a contact or prick sensation which is not painful subjectively. In this the pain system appears to differ from those of warm, cool and touch. Its normal stimuli are mechanical distortion of the skin, and a more intense change of temperature than is required for temperature sense. It is peculiarly liable to increase of irritability under the influence of inflammation. The modality of touch then includes all other senses now known to be aroused by simple stimulation of discrete skin loci or nerve endings, a contact sense normally activated only by mechanical distortion.

A workable classification of cutaneous sensation thus requires a set of criteria which satisfy various conditions; recognition subjectively of discrete and characteristic sensations; specific nerve mechanisms in the periphery; and characteristic and preferential forms of stimulation. A further examination of the characteristics of the pathways between periphery and brain for each sense insofar as at present known may throw further light on the processes by which any stimulus pattern induces its sensory response. This would seem to be an important if neglected aspect of sensation. It will appear that in many respects the whole pathway for a given sense is characteristic of that modality, in the sense that its mode of action, synaptic connections and structure contribute to the pattern of nerve impulses in which the effects of the peripheral stimulus are recorded at the central level.

One of the chief supports for the concept of modalities has been the presumed specificity of skin endings for different forms of energy. Yet von Frey (1896) recognized that the relative thresholds for pain and for touch depend on the area of the contact of the stimulating object; and the structure of the skin itself, and the position of the nerve endings within it (Watterston, 1933; Bishop, 1944), probably play as important a part in selective stimulation as does the selective irritability of the nerve endings. Hardy et al. (1939) find radiant heat a most effective stimulus to pain endings; paradoxical stimulation of cold spots is a familiar phenomenon; all the sense organs except warm spots are excitable elec-

trically (Bishop, 1943) and with relative thresholds for touch and for pain opposite to those disclosed by mechanical stimulation; and Nafe (1942) considers changes in vessel size the immediate stimulus in the response of temperature endings. This author emphasizes the possibility that the nerve endings themselves may be non-discriminative as to forms of energy applied, a possibility previously considered by Bronk (1934). What degree of specificity nerve terminals themselves are capable of is in doubt, and the overall selectivity may best be considered as a function of many factors, position in the skin, encapsulation by non-nervous envelopes, including the dermal tissue itself as a "capsule" or medium through which external agents affect "free" nerve terminals; density or stiffness of overlying epidermis, affected by sweating, etc., and not least, the way in which the energy is applied in customary experience. When variations in these factors are minimized or controlled experimentally the sensation elicited is liable to depart from the casual experience which determines our concepts of sensations, and further complications arise. A certain amount of training of the subject is in fact necessary to obtain reliable data under many experimental conditions. It can then be said that only by taking the "sense organ" as a complex of the sensory nerve terminal, its surroundings in the skin, and the circumstances under which it is usually stimulated, does a high degree of specificity with respect to energy selection result.

The pathway to the higher centers. Between the sense organ sensitive to a specific type of energy and the area of the brain where its sensory message is projected extends the fiber pathway, interrupted at various levels of synaptic junctions. This pathway may be characterized in several ways, other than by its central and peripheral terminations. One is the size of fiber of which it consists, at least in peripheral nerves, together with the fiber properties which go with size. A second is the segregation of afferents into functional groups in the spinal cord. A third is the type of response other than sensation,—reflexes for instance, —induced by peripheral stimulation. A fourth may be the degree of interconnection between parallel pathways of a given sense at synaptic regions, such as would permit spatial summation, for example. In a sense the reverse of this, point discrimination or recognition of position on the body surface, might be a function of the degree of separateness of individual pathways, although to what extent subcortical neurones are involved in this, and to what extent it depends on interpretation of stimulus pattern at the higher centers is unknown. Tower (1943) has offered a theory of localization involving central recognition of the pattern of peripheral ending activation, rather than a simple one-tract pathway from periphery to center, and Lorente de Nó's concept of "partially shifted overlap" (1934), consistent with many observations on the structure and functioning of neurones, presents a picture of a given axon's terminals impinging on the dendrites of several neurones, the selection of the post-synaptic element to be activated depending on the pattern of activity of the pre-synaptics. In the geniculate nucleus of the optic tract a sensory pathway capable surely of the highest discrimination shows a pre- to post-synaptic relation corresponding to this concept (O'Leary, 1940) and certainly not to a one-to-one relation. There is no reason

to believe that pre-to-post connections at the spinal cord level are any more unique, and the differences between the various senses with respect to discrimination of locus on the one hand and spatial summation on the other have at present no final experimental explanation.

The constituent fibers of a skin nerve when classified according to size fall into several groups whose maxima are distinct but whose margins are not (Gasser and Erlanger, 1927). Since threshold as well as conduction rate vary with size, the action currents after conduction segregate into corresponding waves, one for each size group, and the strength of stimulus can be so manipulated as to stimulate fibers only above a selected size. Block by pressure affects large fibers preferentially, and a combination of pressure block and limited stimulus strength will serve to activate a selected band of fiber sizes by eliminating from stimulation both larger and smaller fibers than those selected.

Both in unanesthetized animals and in man (Heinbecker, O'Leary and Bishop, 1933; Heinbecker, Bishop and O'Leary, 1933) procedures of selective activation of skin nerve branches demonstrate that the larger fibers from the skin serve touch only, at any number and frequency of stimulation, while a fairly distinct group of medium-sized fibers (the delta group, 6 to 3 mu. in diameter) include pain fibers. One cannot say from this evidence that smaller fibers, in the delta range for instance, do not carry touch or pressure, and in fact there is evidence that they do; and to be sure the fact that two kinds of touch are recognizable in the skin, that from hair follicles and that from areas of skin between them, leaves the possibility at least that even within the one modality of touch a differentiation with respect to size exists. Other evidence, from stimulation of individual pain and touch spots of the skin, indicates at any rate that in the delta range also touch fibers if present do not mediate pain. Just below the pain size and probably overlapping its distribution in the same group are fibers serving temperature senses (Heinbecker, Bishop and O'Leary, 1934). Still smaller fibers, the C or non-myelinated group, can be demonstrated also to cause pain on stimulation (Clark, Hughes and Gasser, 1935). What other senses this group may mediate is not known at present.

Not all the fibers in each size range are known to be sure to serve these respective modalities, or even to serve sensation at all; further the margins of the size groups are not sharply discrete, nor do the experimental procedures employable separate too precisely according to size. However the demonstration that fibers above a certain size and threshold carry touch but no pain, that fibers just below this in size carry the pricking pain of the skin, and that the largest pain fibers are approximately at the head of the delta group, can be made clearly in conscious human subjects (Heinbecker, Bishop and O'Leary, 1934). Further, differential anesthesia of nerves in man (Heinbecker, Bishop and O'Leary, 1933) indicates that the temperature senses are mediated by fibers in the delta group in general smaller than, but overlapping the pain fiber size. Temperature has not been excluded from the C group. Not only the exclusion of pain from overstimulation of touch fibers, but the convenient and obvious activation of pain endings by heat (Hardy et al., l. c.) suggests that "hot" pain is also mediated by pain fibers and

not by temperature fibers; that is hot or burning pain is a composite of warmth and pain. (Pathological or "burning pain" of the thalamic syndrome is not necessarily so accounted for.) Pain from excessive cold may be not a skin sense, but arise from activation of deeper pain endings, or from both superficial and deep (Wolff and Hardy, 1943); there is no evidence that it arises from excessive stimulation of fibers mediating cold itself. It is perhaps noteworthy that pain from muscle nerves and from stimulation of the sympathetic trunks (visceral pain?) in animals arises also from activation of the same sized fibers as in the skin.

On the basis of animal experiments this concept has been questioned, particularly that pain fibers were confined to the delta and C groups (Gasser, 1943, and others there cited). For instance, stimuli known to be injurious or to arouse pain activate fibers whose conduction rates extend over the whole size range, although it would seem not necessarily to follow that all fibers so stimulated contribute to the pain, without a more discriminating test of their separate functions. Also, the fact that pain is mediated by two groups so far apart as the delta and C might raise doubts as to the significance of size. Nafe (l. c.) arguing for a theory of non-specificity of nerve fibers cavalierly dismisses the idea of specific size group correlation with the statement that the attempt to identify pain fibers by size has failed, and decides that the overstimulation of touch mechanisms causes pain. An enthusiastic reinterpretation of the evidence on human nerve stimulation permits him to draw conclusions the diametric opposite of those of the experimenters he cites (Heinbecker, Bishop and O'Leary, 1933). Other discrepancies reported in the literature on fiber size and function may serve as the basis for legitimate doubts as to the precision of the correlation until these discrepancies shall have been disposed of. On the other hand, the absence of an alpha group in skin nerves and its presence in nerves to muscle provides the experimenter with an ace cold, in that no cutaneous sense can be mediated by fibers as large as the group that carries impulses for the proprioceptive sense of muscle. Stated with due qualifications the conclusions here reached are not irreconcilable with the position of Gasser (l. c.) sometimes cited without his own qualifying statements as the extreme proponent of the opposite view. Beyond a considerable body of accepted findings, divergent inferences as to an ultimate solution are as proper as they are tentative.

From such evidence then it is permissible to infer that the pathways of sense, at least in their course through peripheral nerves, are indeed mediated by fibers of characteristic size ranges, though the ranges for the different modalities overlap at their boundaries. No conclusive reason can be assigned for this distribution; perhaps the higher conduction rate of large fibers makes them more suitable for the more discriminatory senses; touch in the skin and the proprioceptive senses from muscle, etc. Or fast impulses may set the stage for slower ones at synaptic levels (Gasser, 1935). The occurrence of two bands, delta and C fibers, in the nerves' size spectrum which mediate pain may prove to be the salubrious exception, and mean that two varieties of pain are mediated, for which there is some evidence.

It was noted by Goldscheider (1881) that response to pinprick seemed to be

double, an immediate prick followed by a slightly delayed aching pain. Von Frey in repeating these experiments (l. c.) claimed that only the delayed pain arose from stimulation of pain spots in the skin in certain loci where touch was not aroused, and he diagnosed the first sensation as touch, interpreted as pain when followed by the second phase of the double response. The present writer confirms von Frey, and further confirms Lewis and Pochin (1937) that by electrical stimulation of pain spots only one sensation is induced; touch and presumably the pain from C fibers requiring a higher strength of stimulation by electricity than delta pain (Bishop, 1943). However, a double pain sensation has been reported repeatedly and Lewis and Pochin assign the second phase to the slower conduction of impulses in C fibers, assuming both to be activated at the same instant. The change in the character of sensation as the pain from a severe prick develops, and measurements of the reaction times from points at different distances from the cord have led to the suggestion that ordinary prick is mediated by the faster delta fibers, aching pain by the slower. The same aching pain can, however, be elicited as the sensory effect of repetitive electrical stimulation, at a strength such that one shock alone calls forth only bright prick. Unfortunately, all the sensory fibers lose their myelin sheaths as they approach the epithelium, preventing histological identification of point innervation in this respect. Another possibility exists, that non-myelinated fibers serve that pain which is difficult to elicit in normal subcutaneous or visceral tissues, but which is rendered extremely irritable in inflammatory reactions. Lewis and Pochin point out that a non-painful interval between early and delayed pain must indicate a discontinuity of pain fiber distribution between myelinated and nonmyelinated pain groups (larger delta and C), consistent with the known interval in conduction rate.

Until less inferential conclusions can be drawn from more critical evidence, the *differences* in *sensory responses* to delta and to C fibers cannot be said to have been unequivocally demonstrated. Their reflexes do differ significantly (see below). In view of the necessity that C pain must be delayed compared to delta pain if both are elicited at once, it may be suggested that there are two phenomena here confused; an *apparent* double pain when touch and delta pain fibers are stimulated together, and a true double pain from delta and C simultaneous activation; and that their respective elicitation may be a function of intensity of stimulation.

Sensory pathways at the cord level segregate roughly according to modality, and at least on a functional basis. Those mediating touch discrimination pass with the proprioceptive discriminative paths up the dorsal column to synapse in the medulla, and their reflex collaterals synapse chiefly with cells in the dorsal and lateral margins of the grey horn. Pain and temperature fibers enter the tract of Lissauer and leave it near the level of entry to synapse in the substantia gelatinosa. From here the post-synaptic fibers cross to the contralateral spinothalamic tract where temperature, pain and a residual touch localization further segregate dorsal to ventral in the order named. The segregation of the dorsal column discriminatory fibers from the rest at the root entry zone into the medial and lateral divisions of Ranson (1913) separates the large and small fibers of the sensory components of the spinal nerves, the lateral bundle carrying the bulk of the small

myelinated and non-myelinated, to make the functional cord segregation thoroughly consistent with the assignment of modalities to their respective fiber sizes in the peripheral nerves. Beyond the first synapses the sizes of the post-synaptic fibers have not been measured, and admixture with the ascending and descending pathways serving other functions might render this difficult. However, the velocities of the dorsal column continuations of the larger fibers of the saphenous are slow and of smaller size than in the peripheral nerves (Gasser, 1943) which suggests that the value of large size and fast conduction pertains to cord functions rather than to those at higher levels.

One can only speculate as to the significance of this segregation at the cord level, although its correspondence with the classification into modalities on other grounds offers convincing support for the significance of that concept. At the pons and midbrain level these separated pathways, together with corresponding fibers from the cranial nerves which have segregated similarly, rejoin to form the sensory lemniscus passing to the thalamus. So far as is known, touch pain and temperature do not have discrete projection areas in the cortex, although they may have in the thalamus. The discriminatory sensory components are the ones making connection by collaterals with the cerebellum, via cells of the cord adjacent to the dorsal columns, and this may be a factor involved in their spinal segregation. From the facts, however, that the segregation takes place at the level of entry, and that reflex connections are made near that level, and that the reflex responses to stimulation of different modalities are quite different, one might infer that the cord segregation served primarily cord functions, and a search for its rationale might be made in those aspects of cord organization more primitive than is sensory projection to the higher centers. This would be consistent with the view that higher pathways for pain, touch discrimination, cerebellar projection, etc., were added at different phylogenetic stages to a fundamental cord pattern of afferents serving basically spinal cord activity.

Differences in reflex effects of stimulating fibers of different size, and hence of modality, seem to support this view that the spinal segregation is essentially spinal in significance; a view with which the reduction in size of large fibers after they send collaterals to the cord neurones is consistent. Whatever may be the utility of consciousness in organizing body activity, its increasing dominance over body reactions must have utilized the pattern of organization of animals with less elaborate cortices. The function of the brain as a physiological device is to integrate the activities of the lower centers for more complex or more "purposeful" activity. The addition of "voluntary" activity on the motor side, and of conscious sensation on the afferent side to provide a motivation for this activity, is only a further elaboration of the fundamental plan of activation of motor organs in response to stimulation of sensory endings by way of a central synaptic region of the cord segment. From this point of view sensation itself is an incidental if effective refinement, the apical efflorescence on the afferent tree whose lower branches still bear most of the fruit. The preoccupation of the most self-conscious of animals with their own emotions to the extent of narcissism has unhappily led to the isolation of "sensory" physiology from the field of response to stimulation.

While it is true that in certain respects the ultimate as well as the most sensitive and significant criterion available to the physiologist is the response in consciousness of the human subject, this event in consciousness is only one of a group of response activities in general consistent with each other and fundamentally related. It is perhaps not beyond the bounds of rational analysis to speak of modalities of reflex, as of sensory activity.

This point of view has not dominated the field of reflex physiology, with the result that evidence to elaborate it is sketchy and often incidental. It is generally recognized, however, that pain in particular is peculiarly reflexogenic to muscles, a realization facilitated no doubt by the fact that the anesthetic employed in many animal experiments suppresses effectively the reflex response to any stimulus less drastic. Dilatation of the pupil, respiratory and circulatory increases, even more than skeletal muscle contraction are specific enough to be taken conventionally as signs of painful stimulation even under anesthesia. Many reflexes from skin afferents are called forth by touch, and can be demonstrated by electrical stimulation of afferent nerves in lightly anesthetized animals at a strength of shock which the nerve record shows to involve only fibers larger than delta (Bishop and Heinbecker, 1935). Not too much is known concerning reflexes to temperature stimuli below threshold for pain, but the familiar changes in skin blood supply and sweating are obviously reflex. Stimulation of delta and of C fibers of the saphenous separated by differential block (Bishop and Heinbecker, l. c.) gives a suggestive picture of different patterns of reflex responses. With slowly repetitive shocks to delta fibers, each stimulus calls forth a twitch of the gastrocnemius, each twitch followed by an after-discharge if frequency or strength is sufficient. This is in effect a "flicker" phenomenon, and as the effectiveness of stimulation is increased, fusion of individual responses takes place by increase of after-discharge to a sustained muscle contraction. When C fibers only are stimulated, no twitches are obtained, the response to single effective stimuli being a scattered discharge of muscle elements. Fusion to a steady contraction takes place at a much lower frequency, and with no abrupt contractions even approaching a twitch. Granting that temporal dispersion in the slowly conducting C fibers may play a part in this phenomenon, one is tempted to compare these two muscle reflex patterns to the pattern of central activity that must underly prick and ache, respectively.

To some extent, then, the excitation of each modality of sense also excites reflexes equally characteristic of that modality, and equally appropriate for adjusting the animal to a changing environment. That two kinds of pain mechanisms even, delta and C, have different characteristics is supported by the effects of anesthesia on their central connections.

When dial is employed, for instance, stimulation of delta pain is relatively ineffective as compared to stimulation of C fibers (Clark, Hughes and Gasser, l. c.) judged by reflex responses in cats. When ether is administered C fiber reflex responses are entirely suppressed at an anesthetic level permitting brisk muscle reflexes to delta stimulation (Bishop and Heinbecker, l. c.), while in unanesthetized dogs both systems obviously mediate pain. Whatever the central mech-

anisms that differentiate them in this respect, that difference in mechanism is presumably effective in other respects; for instance, it might be expected to be effective at the sensory level. The point has not been sufficiently investigated in conscious human subjects.

While a fairly exact knowledge of the course of sensory fibers through the cord and brainstem has been built up, encouraged by its clinical utility in diagnosing cord lesions, a corresponding knowledge of the organization, behavior and connections other than sensory at the afferent secondary neurone level of the cord is still poorly developed. The tracing of cord pathways and analysis of cord activity by means of the electrical record promises to fill this hiatus. Numerous recent papers bearing more or less directly on this subject present evidence difficult to summarize briefly, and still raise more problems than they settle.

The central projection of cutaneous sense is to the somesthetic area, but no projection of pain pathways is reported to this area by Penfield and Boldrey (1937). Adrian (1941) failed to record responses to either pain or temperature stimuli from the cortex in experiments where touch caused elaborate potential waves. This is consistent with the belief of Head and Holmes (12) that pain is registered at the thalamus. Heinbecker and Bartley (1940), however, recorded discrete responses from the cortex to stimulation of the A, delta and C fibers of the saphenous with latencies about proportional to conduction rates of these fibers in peripheral nerve. Ether depressed the later waves at low concentration, a light surgical anesthesia eliminating responses to all but the largest fibers (touch). The accuracy of discrimination in sensation of both pain and temperature would seem to demand that the discriminating organ par excellence must take part in it, but nobody has demonstrated that even touch discrimination takes place in the post-central gyrus, which is after all only the first point of arrival of touch stimuli at the cortical level, and no one would claim the sensory projection area to be the seat of any sensation. Pain pathways are readily conceivable from the thalamus to those various cortical regions other than somesthetic involved jointly in the complex process of perception, but they have not been demonstrated specifically.

On the other hand, the idea of either thalamus or cortex being *the* locus of so complex a function as sensory perception is losing its appeal. The fact that the decorticate animal shows on stimulation even exaggerated emotional reaction, so-called sham rage, hardly indicates that subcortical organs are the seat of pain, both because such protest reactions can be set off by non-painful touch, and because even to painful stimuli the behavior indicates not perception of sensation, but rather an over-response lacking the degree of purposive control contributed by the intact cortex. There are, however, reports of pain on stimulating the somesthetic area of the cortex at brain operation, and following various cortical damage (Michelson, 1943) suggesting that some part of the pain sensory pathway was involved. At the thalamic level the afferent pathways redistribute their sensory fibers over regions corresponding to their regions of origin in the skin, not so far as known by modality, but rather projecting a spatial body pattern on the thalamus. Disturbance of pain sensibility is considered clinically to be diagnostic of thalamic injury particularly when the so-called thalamic syndrome is

exhibited; an exaggerated sensation of "burning pain" in response to stimuli which would normally seem less severe and less objectionable.

Other situations in which overaction of the pain sense occurs might raise the question whether thalamic injury induced this effect by any other means than the interruption or irritation of some of the pain pathways. A similar increase in effectiveness of peripheral stimuli takes place during the re-innervation of an area whose nerve supply has been cut, a phenomenon which Rivers and Head (1908) took to be evidence of the pain they called protopathic. Irritative lesions along nerve trunks may occasion similar increases in the effectiveness of peripheral stimulation, especially when some of the fibers are blocked as well as irritated. In a case examined by the writer (unpublished) after operation for focal epilepsy, in which removal of a portion of the sensorimotor cortex had partially paralyzed and partially anesthetized the left hand and lower arm, this overaction to pain stimuli was pronounced. Examination of the pain spots over the affected area revealed that about half of the usual number, or of the number on the opposite side, were sensitive to prick, but stimulation of one of these, or of a nerve twig under the skin, gave the exaggerated effect. The condition seemed to be that the central representation of about half the pain spots had been destroyed by removal of cortical material only, in the hand area for touch, leaving a situation comparable to partial regeneration of a peripheral supply. This patient, medically trained and superintendent of a hospital, finally arrived at the conclusion that a given painful stimulus to the affected hand could be estimated quantitatively as accurately as on the other hand, but that its objectionable character was greatly enhanced. He volunteered the suggestion that part of the repugnance to stimulation of that hand arose from apprehension, due to imperfect recognition of a customary pattern; he often fumbled a cigar in that hand, for instance, so that the lighted end burned his fingers.

Taking such evidence, it seems possible that over-estimation of the painful character of a stimulus in the thalamic syndrome and similar states arises from a defective pattern of central representation, caused perhaps by a reduction in the density of pain spots represented in the cortex, due to interruption of some of their pathways anywhere along the course to the brain, rather than by injury to some specific center at which pain is perceived. The effect appears to be principally on one aspect of sensation, which Hardy, Wolff and Goodell (1943) have termed the *reaction* to pain, and which has otherwise been referred to as the dolorous affect, as compared to the *perception* of pain or other senses as a discriminatory process. The fact that pain is affected by such lesions in a more noticeable degree than are the other skin senses is consistent with the fact that pain is the one sensation accompanied to an extreme degree by this emotional or affective reaction. The nature of the change in response in sensation with defect in the normal pattern cannot at the present stage be interpreted on a physiological basis.

The unique character of pain. As that sense particularly designed to warn of danger, the central perception of pain has associated with it an unpleasantness as such, which builds up and "overflows" into an emotional tone of aversion and

protest. As noted above, the type of reflex and the character of reflex response induced by "painful" stimuli, even in unconscious animals, has something analogous to this, rendering the whole pathway and its connections at all levels characteristically a unit in promoting avoidance of injury. This is not to say that other senses are incapable of inducing a similar effect, or that activation of pain pathways always and necessarily does so. A comparison of the attitude of the subject undergoing pain stimulation as an experimental procedure with that of the sick and anxious patient whose pain is mysterious, unpredictable and of unknown causation, not to mention the factor of persistence of the pain, indicates that in casual experience the *reaction* to pain may be of more significance to the animal than its mere perception.

On the other hand the threshold activation of pain endings gives rise only to an almost unnoticeable contact sense (Bishop, 1943) turning to a sharp and bright prick with increased intensity of stimulus, which experimental subjects, well or sick, do not usually interpret as painful. Above this level pain is reported at various stages, depending on many factors, not the least of which is the mental attitude of the subject. Repeatedly subjects receiving touch stimulation only, from electrical shocks to nerves, have reported it painful, only to revise the report when pain threshold was reached, stating that the abnormal touch sensation was unpleasant and *threatened* pain. In stimulating nerve twigs in conscious subjects, the stable and reproducible division point is between touch and prick, not between touch and pain.

Hardy, Wolff and Goodell (1940) using radiant heat as a stimulus find that taking an initial prickling sensation as threshold, this level of perception is remarkably constant both from day to day and under the action of various factors. This is, of course, in an experimental standard state where prick need be differentiated from warmth alone. In other words, the brain can perceive the results of peripheral stimuli and discriminate between them with relative constancy when nothing else sufficiently complicates the situation. These are not the circumstances under which one usually experiences pain, or even the complex non-painful sensations to which the activation of pain endings among others contributes. Inflammation or injury to the skin lowers the peripheral threshold (Schumacher, 1943). Many analgesic drugs raise the threshold of perception (Wolff, Hardy and Goodell, 1941) as does most brain damage. That is, such drugs either decrease the efficiency of response of peripheral endings or decrease the acuity of the brain in perceiving the results of their activation. In addition, many drugs decrease more remarkably the *reaction* to pain, or the affect; a result also of distraction, freedom from apprehension, etc. (Wolff and Goodell, 1943). It is not clear in view of the obvious central effect of drugs whether they have any effect on the periphery in ordinary analgesic dosage, nor is it always clear whether the increased perceptual threshold under drugs, etc., is in effect a result of changed mental attitude; lack of attention, interest, or less careful discrimination, for instance, which to be sure are themselves factors in the complex act of perception itself.

In one case this point of differential effect on two phases of the same modality

has been scrutinized. In a study of the effect of amphetamine on tooth pain excited by electrical impulses through a filling Burrill et al. (1944) examined the two thresholds discretely, of first sensation, and of report of pain. No tactile sensation is aroused from the tooth pulp; below actual pain, a vague sensation is elicited that is non-painful but is reasonably assigned to the activation of pain endings. This is comparable to the fact that weak stimulation of the cornea which also has only pain endings is perceived as an unpleasant but not painful contact (Tower, 1940) and the same holds for stimulation of skin pain spots below threshold for either prick or ordinary touch (Bishop, 1943). Likewise "pre-headache" sensations (Schumacher and Wolff, 1941) precede overt headache induced by histamine, and clinically, vague distress below the level of pain is encountered from the viscera that presumably arises from activation of pain endings below threshold for frank pain.

In this respect tooth sensation is a favorable choice for investigation, because no such sensation as non-painful prick of the skin intervenes between the vague recognition of stimulation below pain and pain itself; the first painful twinge is obviously of the identical character as more of the same. Under these circumstances 15 mgm. of amphetamine raises the threshold to pain but not to a first tooth sensation; more raises both thresholds. The effect is interpreted as being entirely central, that is, under this analgesic it takes more sensation to be estimated as painful.

From this general picture, *three* thresholds may be recognized in the activation of pain endings generally; that for first sensation, that for a sensation recognized as sharp, pricking or painful in perception but without emotional protest, and that for aversion to pain, Wolff's *reaction*. The transition from prick to pain, where prick occurs, is not sharp; prick merges gradually into pain. Even these do not include the absolute threshold for sensory nerve fiber activation, which is the arousal of the first impulse in one nerve fiber, and which is presumably not perceivable at all. No one knows how many impulses are required to break through the various synaptic obstacles to give a minimal response at higher levels, but it is certainly more than one.

There can be differentiated then at successive stages of intensity of stimulation of a given set of "pain" endings, first a nondescript threshold contact sense, then prick, merging gradually into pain which under experimental conditions at least is recognized as "painful", i.e., is perceived, without emotional protest; followed finally by a degree of this same painful sensation which becomes a definitely objectionable experience, and induces an emotional reaction of protest and other physiological evidences of an avoidance or protective reaction. It is presumably this last degree of stimulation that corresponds to the "pain" reflex responses of animals. The fact demonstrable in conscious human subjects that not only analgesics and anesthetics but emotional, and particularly anxious and apprehensive states affect the threshold of this protest reaction more than they do the threshold of perception, is the chief physiological reason for considering pain reflexes to be equivocal indices of pain intensity in animals. The extent to which this reaction threshold differs from the *threshold stimulation of the pain modality*

in animals may depend on the condition of the animal much as do conscious pain sensations in the human subject.

A feature of this situation not sufficiently considered is the existence of non-painful "pain", which in casual experience must involve the contribution to many complex and non-painful sensations of components assignable technically to the pain modality. Aside from the fact that persistent electrical stimulation of pain spots below threshold for pain induces itch, the scratch which follows, not to say pursues the itch, whatever else it accomplishes must stimulate also pain endings, in the paradoxical suppression of a phenomenon in its own modality. Once the threshold, not of pain reaction, but of the perception of the minimal reaction of pain endings is recognized, the low threshold of this response permits participation of pain endings in any contact more drastic than the blandest touch. Brushing the scalp, mopping the brow, vigorous rubbing and especially mild scraping of any surface and firm manipulation of tools, instruments, rough surfaces, etc., must constantly activate pain endings, which participate in discrimination, information and control of environmental objects far short of that activity inducing the protective reaction or the apprehensive emotion usually associated with pain.

The activity of skin sensory endings. Following the classical experiments of Blix, Goldscheider and von Frey, the study of skin endings was for a period directed largely to the location and histological identification of specific sense organs. A study by von Frey (1896) of the difference in the character of skin distortion which activated preferentially touch and pain endings led him to assign touch stimulation to simple pressure distortion, pain stimulation to hydrostatic pressure differentials due secondarily to distortion. A mathematical formulation of the supposed stresses accounted for the differences in actual thresholds. Many subsequent experiments varying pressure and area of the stimulus object added little to the understanding of the essential character of sensory excitation. The finding of specific selectivity of peripheral endings was complemented by Mueller's doctrine of specific nerve energies: that however a sensory pathway was excited, its activity aroused the sensation characteristic of its central apparatus.

With the development of electrical recording of nerve impulses, and Adrian and Bronk's (1928) technique for recording from single sensory fibers, a third factor became apparent; different sense organs activate their nerve fibers to different patterns of responses, each appropriate to the function which the given structure serves at the reflex or sensory level. The chief differences in pattern involved may be summarized under the terms *adaptation* and *after-discharge;* a falling off of frequency under a constant stimulus, and a continuation of activity after the stimulus is removed. They seem to be to some degree independent variables of sense organs.

How these properties are determined at sensory endings is unknown. With the exception of the simple retina of Limulus (Hartline and Graham, 1932) and by analogy perhaps the retinogram of higher forms, no response of any sense organ has been directly recorded, at least if one presumes that the sensory ending differs in action from the fiber it activates. Cogent arguments have been presented that

it does differ (Bronk, 1936). This is demonstrable in the Limulus eye in that a long-lasting slow wave from the retinal cell body is accompanied by a train of typically brief impulses in the nerve fiber activated. Otherwise the only information we have is derived from the temporal pattern of impulses in a sensory fiber when its ending is stimulated. For instance, since responses in tactile fibers may follow the stimulus frequency applied to the skin to as high a rate as the fiber itself can be stimulated, Adrian, Cattell and Hoagland (1931) conclude that the tactile ending must have a refractory period as brief as that of its nerve fiber. But whether this really means that the sense ending itself responds repeatedly at such a rate, or merely that each stimulus raises its persisting excitation to the point where it discharges its fiber, is not so certain. Other evidence might suggest that some sense organs at least do not respond in an all-or-none manner, but rather undergo a more persistent and graded excitation which activates the fiber periodically as a steady galvanic current does, at a frequency depending on the strength of the stimulus and decreasing with adaptation of the fiber to the stimulus.

The nature of adaptation, of sense organ, nerve fiber or central synapse, is not clear; it is exemplified in the relative ineffectiveness of slowly-increasing stimuli, and in the fact that a change in intensity of any stimulus is more effective than a steady state, and appears to be a property of all irritable tissues to all stimuli. However, adaptation in excitable tissues can be materially modified. In nerve the deprivation of calcium reduces it, at the same time increasing the irritability, until the fiber may respond spontaneously. Treated with oxalate, frog touch endings, normally rapidly adapting, adapt more slowly, like normal pressure endings (Talaat, 1933) and finally become spontaneously and repetitively active, when adaptation is either abolished or effectively masked. It is possible that the effect of inflammation on pain response may involve changes in adaptation. On the other hand as the "sense organ" should perhaps be considered as the whole complex of tissue surrounding the nerve ending which conditions the transformation of applied energy into nerve impulses, (supra) the mechanical properties of tissues undergoing distortion or physical stress will introduce a factor appearing as adaptation. For instance, Nafe and Wagoner (1941) find that stimulation of contact endings parallels motion of the tissue under the stimulus, and conclude that adaptation consists of the gradual decrease of motion as resistance to further deformation balances the distorting force. Adaptation is thus assigned to the non-nervous tissue surrounding the nerve ending. Although it is common experience that adaptation to adequate stimuli may not be complete even after an indefinite time for mechanical adjustment to applied energy, and that adaptation takes place to stimuli involving no motion in tissues, it is still possible that any factor inducing a decrement in effectiveness either of stimulus intensity or of fiber irritability will appear as adaptation under otherwise constant conditions.

In terms of nerve fiber response, the characteristic pattern of a number of sense organs has been described, and Bronk (1934) has discussed the correspondence of these patterns to the type of sensation or reflex response on the one hand, and to utility in adjustment to the environment on the other. For instance, stretch

receptors of the lungs (Adrian, 1933) and of the carotid sinus (Bronk and Stella, 1935) adapt only slightly to persistent stretch, and show slight after-discharge. Muscle stretch receptors (Bronk, 1929) adapt but little, and these and the receptors of the carotid sinus actually exhibit what might be called negative after-discharge, a silent period accompanying reduction of tension, which is followed by a reduced rate corresponding to the new tension. Pain receptors (Tower, 1940; Hogg, 1935) adapt but little or even show the opposite phenomenon, inertia, indicated by a progressively increasing response during constant stimulation, and they show a persistent after-discharge. Tactile endings adapt with extreme rapidity and completeness (Adrian, Cattell and Hoagland, 1931) without any after-discharge, although removal of the steady stimulus acts as a second stimulus. Pressure endings in the skin or subcutaneous tissues adapt much less than do touch endings, and their after-discharge has not been specifically reported. In the optic (Hartline, 1938) and auditory nerves (Galambos, 1944) an array of fibers is found which if classified in terms of adaptation and after-discharge not only covers the gamut of sense-organ combinations, but raises the question of the significance of these terms. Some fibers show prompt adaptation, with no response at cessation of stimulation; some respond at cessation as well; others show but slight adaptation, or show a relatively silent period following an initial burst, followed again by a higher discharge; still others fail to respond during a stimulus but respond only at its termination. Some of these phenomena may be complicated by the fact that the impulses are recorded postsynaptically to the sensory ganglion cells, which suggests that further investigation at the cord level of cutaneous sensory pathways might be profitable, in elucidating the modification of impulse pattern along the pathway from periphery to cortex. In other words, adaptation and after-discharge may occur centrally to further modify the message to the center initiated by a given peripheral stimulus, compromising somewhat the interpretation of sense organ action from the report of the course of sensory experience.

Still these differences in pattern of response of sensory endings, plus the modification of pattern at synaptic regions, must largely determine the form of the message received at the conscious level from a given pattern of stimulation. Granted that the sense organ is initially selective with respect to different forms of energy, to what extent is its response, however stimulated, also characteristic of the modality it serves?

For pain and contact senses, obvious correlations exist between sense organ action and the type of information to be conveyed to the center (Adrian, 1928; Bronk, 1934). A threat of danger mediated by pain endings would be misleading if the sense organ responding ceased to act before the stimulus were disposed of; but the vibratory sense served by touch would be ineffective unless its sensory receptor were "dead beat", that is, promptly adaptive and without after-discharge. Pressure and tension endings, non-adaptive and extremely sensitive to changes in stress in terms of frequency of response, follow persistently and flexibly any changes in their environment that affect them at all. Concerning the temperature senses less is known in this respect, since animal experiments so

far dealing with these probably included responses of pain endings as well as of temperature endings to heat, but it is safe to anticipate a similar correlation.

It may be noted that within the one "touch" modality, two types of endings, selective for two different aspects of contact, touch and pressure, transmit to consciousness two characteristically different types of pattern from a given pattern of stimulus. Within the pain modality a further differentiation is evident, in which one type of ending induces different qualities of sensation by means of impulse patterns imposed on it by different stimuli. While the different varieties of touch sensation, simple contact, moving contact or rubbing, repetitive contact or vibration, etc., may be readily recognized to have the common factor of bland touch, activation of "pain" endings runs a gamut of sensations from an indifferent contact sense through itch, prick, bright pain and ache (Bishop, 1943). To verify the assignment of all these sensations to one and the same type of ending electrical stimulation is useful for several reasons. First the threshold for pain to electrical shocks is lower than for any other modality of skin sense. Thus, if itch, for instance, is not induced by stimulation of pain endings, at least it is not assignable to touch, pressure or temperature endings. Further the precision with which brief shocks can be manipulated into patterns of repetitive stimuli, and particularly the possibility of applying and removing them instantly, permits a precise knowledge of the applied stimulus for comparison with the response. While it is not feasible to record directly from nerve fibers in conscious subjects, knowledge gained from animal experiments permits the pattern of nerve response to be inferred from the stimulus, to be then compared with the sensation induced. It is then not impossible to infer from the sensory responses to specific patterns of stimulation the more complex patterns of response corresponding to more casual contact with the environment.

Within these conditions, a persistent and annoying itch quite similar to that of an insect bite, is set up by rapid repetitive stimuli, each shock so weak that summation of the effects of several is required to produce any sensation at all. It has a long after-effect, presumably due to the after-discharge characteristic of pain endings, which is promptly allayed by rubbing or scratching. From the type of stimulus which induces itch, one may infer that the relief of itch by scratching is brought about by breaking up the monotony of nerve fiber response inducing the itch, not by abolishing the activity. Itch also follows sharp prick as an after-effect of single stronger shocks. To weaker single shocks, bright prick without itch is elicited; repetitive shocks as slow as 5 per second show pronounced summation, the overall sensation mounting over a period of one-half to one second, punctuated by increasingly sharp stabbing prick. Above this range of strength prick shades gradually into pain, and pain that is received without aversion into pain to which the hardiest subject will finally protest. In these various gradations frequency of repetition and intensity of shocks are equally effective in determining the amount and quality of sensation, indicating the effectiveness of both spatial and temporal summation, and of the summation of the after-discharge of one volley with the response to the next. Adjacent pain spots also summate effectively. In this sequence of sensations, foreknowledge of the

stimulus pattern permits a further qualification of the character of sensory discharge corresponding to each sensation.

The threshold contact sense of the pain modality is merely the minimum of any activation which is able to break through into consciousness, too limited in amount to be discriminated. Itch corresponds to a slow but persisting and uniform frequency of fiber response, which even follows the sharp prick of a single strong shock by reason of the gradual falling off of the after-discharge, approaching a steady frequency. Prick and pricking pain differ from each other only by intensity of stimulation, therefore in number of fibers activated and the frequency of their response, that is, in the total number of responses per unit time arriving at consciousness from a given locus. Aching pain differs from pricking pain apparently by reason of duration, that is, by the degree of central summation over time, of strong stimulation. Its characteristic tone is not induced by one shock, however strong, nor by any number of shocks if weak. Only myelinated or delta pain has so far been analyzed in this respect.

While the pain modality exhibits perhaps the greatest qualitative range of sensation, touch is notoriously discriminative of the character of the stimulus object (smooth, uneven, moving, etc.) even when other modalities can be excluded from stimulation. Skin tickle is apparently a form of touch as itch is of pain. It seems to require an intermittent stimulus, or successive stimulation of adjacent points (movement) and stimulation below a certain intensity. It can be faintly elicited by a hair attached to a tuning fork on one touch spot on the palm of the hand at an intensity below threshold for pain spots, but is increased by moving the vibrating hair across successive touch areas. A wisp of cotton moved lightly over the upper lip can cause excruciating tickle, which is not obtained, however, if the pressure is increased, nor is it obtained by the lightest pressure from a mass of cotton that stimulates a larger area of the lip at once, even by a moving contact. The temperature senses are probably less flexible in this sense, although their complex summation with touch and pain to give wet, hot, clammy, etc., are obvious.

It will be apparent that these gradations in sensation within one modality connote most effective summation, spatial and temporal, at the conscious level of the effects of individual nerve impulses. In contrast to this, Wolff has been quoted in the literature to the effect that pain is not summated. Wolff's careful and definitive statement, however, (1943) is that "*beyond a minimal limit* in the normally innervated skin the *threshold* for pain was dependent alone upon the strength of the stimulus and not on the number of sense organs stimulated. . . although the total distress to the individual is dependent both upon the intensity of the stimulus and the size of the painful area." (Italics ours.) Further he notes that in nerve lesions reducing the density of pain endings even the threshold for pain is increased. This is to state that the *perceptual* threshold of first prick from a few endings is the same as that from many, provided that those few endings are normally concentrated; but that the *reaction* to pain depends upon the number of endings stimulated, as well as upon the number and frequency of discharges from each; that is, it is a function of the total number of impulses

per unit time. Since this *reaction* to pain is to the casual sufferer pain itself, to quote without technical qualification that pain is not summated conveys the opposite of Wolff's findings as well as of his conclusions. Wolff's statements apply further only to spatial summation, and the fact that his threshold response is measured at the end of a significant period of stimulation certainly does not exclude temporal summation, although it does not necessarily indicate it.

That the experience of pain is about the most responsive to summation, especially to temporal summation, of any of the cutaneous senses, at least more so than touch, would epitomize common experience, and experimental stimulation of individual pain spots confirms this. It reveals further a peculiar relationship between summation of intensity or of effect of sensation, and two-point discrimination for pain (Bishop, 44a) in that one tends to exclude the other in consciousness. This is consistent with the fact that pain is both a discriminating and an affective sensation, being localized almost as well as touch, while capable of arousing an emotional effect more vigorous than that of any other sensation. Consciousness appears to be to some degree limited to a choice between discrimination of locus and evaluation of quantity, and that choice is conditioned by peripheral as well as by central factors, that is, by the spatial and temporal pattern of impulses reaching consciousness as determined by the pattern of stimulation.

Discrimination in general, of locus or of intensity, is poor near the threshold of stimulation for any sense. Above threshold of perception, increase of central effect can be a function of at least three variables in the periphery; increase in the number and frequency of impulses from a given sensory ending or group of them corresponding to temporal summation; increase in the number of endings activated within a given modality, comprising spatial summation; and changes in relative degrees of activity of different modalities, or the overall pattern. Increases in total stimulation of any pattern may have both quantitative and qualitative effects. Moreover, since the stimulation of a single ending, or even of endings of a single nerve fiber, is usually accomplished only under highly technical conditions, "pure" sensations, not to say simple stimulus patterns, are the exception in ordinary experience. Slightly above an intensity suitable for touch, pain endings are liable to be stimulated below threshold for pain, and thus contribute to contact sense. Fusion of contact and temperature senses is so usual as easily to escape notice, the sensation being given a special designation as wet, slimy, etc., or the contact aspect is ignored, although it contributes to the quality of the temperature sense. Similarly touch, pressure and muscle sense are compounded in the estimation of weight or force. In these complex patterns each aspect of the stimulus serves as a clue to the interpretation of the others in evaluating the external object or agency furnishing the stimulus.

Aside from the associated emotional effect, the useful perception of sensation involves discrimination of intensity, quality or space. Granted that intensity is obviously a function of frequency or of total number of impulses, and that modality (quality) is a function primarily of specific central connections, however complicated by concomitant circumstances central or peripheral, the question

remains, what aspect of the sensory pattern permits discrimination in space. The naïve inference that spatial localization could be accounted for by direct spatial projection of the periphery on the cortex, with separate point-to-point connections by means of simple end-to-end chains of neurones has met with too many difficulties to be enumerated here, but enough to demand a more complicated explanation, and perhaps an explanation of more than *spatial* perception. It must be granted that some degree of localization of the periphery on the cortex is indicated by the separate projection areas of the special senses, by the representation of successive regions of the body along the central sulcus, and particularly by the reproduction within the visual area of the geography of the retina. Work on the peripheral distribution of sensory endings (reviewed by Tower, 1943) reveals that spatial localization, two-point discrimination, etc., is more precise than can be accounted for by the discrete distribution of separate fiber endings over the body surface, and in fact reveals that wide overlapping takes place in the ramifications of individual fibers. What is known of the complexity of interconnections between parallel fiber pathways at synaptic regions, and which may be inferred to be especially pronounced at the cortical level, adds to the same difficulty. Starting apparently with a basis for crude localization in region-to-region, not to say point-to-point projection of periphery onto center, the brain has elaborated the fineness of its discrimination beyond the limits which the degree of specificity of anatomical connections would permit, and without a corresponding increase in the detail of such projections. Without attacking the problem of how consciousness perceives the pattern of activity of the brain, which must still be left in the realm of psychology, inferences are in order as to what patterns of activity reach the brain, corresponding to the perceptual acts of which it is capable.

Such inferences have been made by Tower (1943) in the form of a theory of spatial localization, similar to an earlier theory of Bethe (1895). Based upon experimental examination of the responses to mechanical stimulation of single pain fibers of the cornea, and justifiably applied to other areas of the body surface and to other modalities, Tower employs three findings; that the neurological unit is one fiber and its ramifications, without anastomosis between fibers; that the ramifications of single fibers overlap widely; and that different regions of the area of ramification of a given fiber show differences in the properties of adaptation, excitability, refractoriness, etc. If then a point stimulus is near threshold intensity, only one fiber may be activated, and even if projection to the cortex were sufficiently discrete to permit of the discrimination of one fiber's responses, the area of distribution of that fiber is too extensive to permit as precise localization as stronger stimuli are found to make possible. If now the stimulus is increased, different fibers with endings near the point will be activated to characteristic patterns of impulse discharge, which depend on the region of each fiber's distribution that coincides with the stimulated point. The central brain must then recognize as a sensory "point" that pattern of different responses in overlapping fibers, and not the whole area reached by branches of one or all of the fibers activated. A given fiber might then be utilized in the identification of

points anywhere within its distribution, its activity contributing to different patterns depending on what other fibers were activated with it, and to what degree. Following nerve block or destruction when the concentration of endings from other nerves at the margin of the denervated area is reduced, localization is less precise because fewer fibers can contribute to an adequate pattern of response. The sensory spot is thus a central interpretation of a peripheral configuration of activation. This local "sensory spot" configuration is presumed to consist of a central region of one fiber's distribution where this fiber is most sensitive and can respond maximally, surrounded by the terminal ramifications of other fibers whose responses will be less vigorous to a given stimulus, in a more or less heterogeneously overlapping peripheral network.

Minor refinements of this idea may be in order as more details become available concerning the distributions of nerve endings in different tissues, without rendering the basic proposal less plausible. In the skin of the arm, for instance, after removal of the papillary layer of dermis containing most of the pain and touch terminals, discrete nerve twigs are found upon stimulation of the cut surface, and can be demonstrated histologically during regeneration of the pain spots (Bishop, 1944b), the center of a pain spot developing over each growing twig. Accounting thus to some extent for the anatomical individuality of pain spots, their physiological individuality is still conditioned by the circumstances noted above, and in fact successive anesthesia of adjacent nerve branches with overlapping distributions reveals that some spots at least are partially blocked from each nerve, in confirmation of the histological observation (Weddell, 1941) that many sensory units are multiply innervated from branches of separate fibers. This again demands that localization must depend on more than the activation of one fiber, and suggests that the anatomical arrangement into peripheral twigs, of whatever fiber source, accomplishes a local concentration of endings suitable for the activation of a pattern of stimulation that can be effectively discriminated.

It is obvious here that the idea back of the term "sensory unit" has been a confused one. Actually we may speak of three types of units in this connection. There is first the anatomical neurone unit, consisting of one fiber and its more or less extensive ramifications, overlapping with but insulated from every other unit in the periphery. Secondly, there is a structural unit consisting of the endings supplied by terminal fiber branches of one nerve twig to an overlying area of skin, the sensory spot, characterized by having a most excitable point at its center and a less sensitive periphery. Third may be specified the physiological unit locus, which is the lower limit of area whose position on the body surface can be identified, or whose discreteness from adjacent units can be recognized in consciousness. The latter unit is variable, not only over different regions of the body but depending on the intensity of stimulation, sequence of stimuli and other factors. The structural unit supplied by one nerve twig and recognizable as a sensory spot is perhaps the lower limit of the physiological unit. The ramification of a single fiber might be considered the lower limit of a sensory spot, when such a structure is found in isolation from overlapping fibers. When

such simple sensory spots are found, as for instance in the margin of a denervated area, spatial discrimination is poor, and intensity perception and affect are abnormally acute. This may be another aspect of the tendency noted above for spatial discrimination and spatial summation of intensity to exclude each other, and its analysis may contribute to an understanding of the "burning pain" of the thalamic syndrome and similar states resulting from impaired innervation.

The phenomena of spatial discrimination may offer a practical approach to a more general understanding of sensory processes. Kellgren has suggested (1939) that pain from various parts of the body might be classified on such a basis. He points out that there are three *layers* of sensation, the skin where discrimination is most precise being the first, below which, in a second layer of deep fascia, periosteum and ligaments, localization is of only fair accuracy, not equal to that of the skin. From all structures below this second layer pain is diffuse, poorly localized, and liable to be referred within a segmental distribution. It is referred, in fact, to those regions of the segmental distribution *in which pain is better localized*, suggesting that the segmental distribution of diffuse pain, i.e., the diffusion of pain reference to other parts of the distribution than its origin, is simply a false or imperfect localization. It will presumably be true that rather than three distinct layers, a gradient or transition exists from surface to viscera, and even from one area of surface to another, and that the ability to discriminate space will be a correlate both of the density of innervation and of the experience of the individual, i.e., of training. At least it can be said that precision of localization varies with the probability of specificity of stimulation, those parts of the body which are exposed to the most localized or definitely patterned stimulation, the hands for instance, being the ones which are capable of the finest discrimination.

This concept would seem to be applicable to other senses than pain, particularly to the contact sense of touch, and especially to the various combinations of senses that give us our casual sensations. What Tower's theory proposes for touch or for pain, that what the cortex recognizes as a unit locus is not a local stimulation merely but a specific pattern that is aroused by stimulation at a given point, is even more applicable to the more complex pattern involving *both* pain and touch, plus other modal varieties to be fused into a discriminating recognition of the character of the stimulating object. To take an example from common experience and universally accessible, a mild itch is less accurately located than a severe one, it is more accurately located on the hand or face than on the back or thighs, and localization is extremely facilitated anywhere by light touch, particularly by successive touch say from a moving finger. The tendency to manipulate or press upon poorly localized deep painful regions arises perhaps from an indefinite impulse to better localize the pain, by adding to it other more critical sensations. Such combined sensations of which pain is a component are more easily bearable than is pain by itself, and the distressing character of visceral pain may be in part assigned to this inability to localize it precisely; that is, to fully evaluate it in total experience.

It may be concluded in fact that only when the activation of one modality

predominates to such a degree as to exclude lesser activities from consciousness, or except where the same process of exclusion is accomplished by deliberate or attracted "attention", or when both factors are brought into play as under most experimental conditions, is one justified in treating any one modality of sensation as a discrete phenomenon. Otherwise consciousness seems to pay much less heed to the concept of modalities than do the physiologists who try to differentiate between sensory processes.

SUMMARY

It appears to the writer then that the following statements are for the present tenable.

1. Both sensation as a central phenomenon, and the activity of sense organs in the periphery, may be conveniently reduced to a common denominator in terms of the pattern of non-specific impulses in the fibers of peripheral nerves.

2. The concept of sensory modality has become a straight-jacket from which the study of sensation may profitably be released, to the extent of relegating this vestment to the status of a loose-fitting generalization.

3. The pathway to the higher sensory centers is interrupted at successive levels of synaptic regions, where two processes can intervene. First, individual fibers can interact to accomplish spatial summation and mutual facilitation. Second, collateral activation can result in reflex or other activity characteristic of the given level.

4. Each class of sense organs activates fibers within size ranges whose maxima at least are significantly characteristic, although the margins of the groups overlap. Touch and light pressure show two maxima, in the beta and delta fiber ranges. Pain shows two, in the delta and C ranges. Warmth has one just below pain in the delta group, and cold has one just below warmth, and temperature cannot be excluded at present from a second representation in the C group. The "waves" of classic nomenclature may in fact be the result of gross summation of such functional groups. The double representation of certain functions in two fiber positions may mean that two aspects of these senses differ in their physiological significance, although giving rise to one sensation in consciousness when aroused undiscriminatively. The utility of fiber size distribution is still a subject for speculation, and is widely so employed.

5. The segregation of sensory pathways on their entrance to the cord seems to be related chiefly to reflex function at the cord level, rather than to sensation at the cortical level.

6. From the point of view of utility to the organism the different reactions at different levels of a given sensory pathway are consistent and directed toward a common end. One may speak of modalities of reflex action as justifiably as of sensation. Sensation is only the final response of such a pathway at the highest level, and is not necessarily its most useful or physiologically significant one.

7. All the sensory pathways find expression at the reflex, thalamic and other levels. The cortical projection area of a sensation is not necessarily the cortical locus of that sensation. The sensory process presumably involves a considerable

amount of cortex and thalamus and possibly of lower centers, and the remainder of crude sense following injury or removal of one region should be looked upon as involving a defect in the normal sensory response, rather than as isolating the focal representation of that sense. The "thalamic syndrome" may be viewed as a defective registration of pain, and similar overaction to painful stimulation may result from defects in the periphery.

8. Pain is unique in the degree to which its arousal overflows into affective or emotional protest, although other senses also share this capacity, which may be accentuated or depressed in abnormal mental conditions. Below a certain intensity, whose level varies both with mental and emotional states and under anesthetics and analgesics, pain may be perceived without significant affective reaction. Below the intensity of stimulation required for pain, activation of pain endings induces sensations qualitatively different from pain, as prick, itch, etc., depending also on the pattern of stimulation. These non-painful "pains" fused with touch and temperature may contribute to the complex sensations of casual experience not recognized as partaking of the character of pain.

9. Sensation is a function of the pattern of responses of the nerve fibers reaching the higher centers. This pattern, perhaps modified at successive synaptic levels, is ultimately determined by the pattern of peripheral stimulation. The different classes of sense organs are not only preferentially susceptible each to its characteristic form of energy (mechanical force, temperature, etc.) but also have properties of their own which affect the pattern of response in their nerve fibers. Touch, pressure and pain for instance show different degrees of adaptation and after-discharge to a steady stimulus, and thus set up characteristic patterns of impulses in the peripheral nerves. This emphasizes the fact that all parts of each sensory pathway are adapted to the useful performance of the particular type of function served.

10. The precision with which sensation may be localized in the body, that is with which sensation may detect which part of the body is being stimulated, varies with the different senses, over different regions of the body, and at different depths below the surface. The capacity to localize accurately may be correlated with the utility of the part in localizing objects in space, as well as with the density of innervation. Sensation in a part incapable of precise localization may be referred to the region within its own segmental distribution better able to localize. "Referred" pain is then simply pain in a region poorly localizable. Since localization over the more sensitive areas of the body is more precise than the rather wide distribution of the endings of single fibers would account for, it is necessary to account for a localizable sensory spot as being determined by the pattern of sensory discharges set up by a point stimulus in any nerve fibers part of whose sensory endings lie within the spot. Consistent with this concept, a weak stimulus activating only part of these endings to an incomplete pattern is poorly localized. In casual experience where more than one "sensation" may be aroused by a given stimulus object, the whole is fused to a complex sensation in which each component serves as a cue to the interpretation of the others, the more complete pattern of activity being again the more accurately localizable.

REFERENCES

ADRIAN, E. D. Afferent impulses in the vagus and their effect on respiration. J. Physiol. **79:** 332, 1933.

Afferent discharges to the cerebral cortex from peripheral sense organs. J. Physiol. **100:** 159, 1941.

ADRIAN, E. D. AND D. W. BRONK. The discharge of impulses in motor nerve fibers. J. Physiol. **66:** 1, 1928.

ADRIAN, E. D., McK. CATTELL AND H. HOAGLAND. Sensory discharges in single cutaneous nerve fibers. J. Physiol. **72:** 377, 1931.

BETHE, A. Die Nervenendingungen in Gaumen und in Zunge des Frosches. Arch. f. Mikr. Anat. **44:** 185, 1895.

BISHOP, G. H. Responses to electrical stimulation of single sensory units of the skin. J. Neurophysiol. **6:** 361, 1943.

The peripheral unit for pain. J. Neurophysiol. **7:** 71, 1944a.

The structural identity of the pain spot in human skin. J. Neurophysiol. **7:** 185, 1944b.

BISHOP, G. H. AND P. HEINBECKER. The afferent function of the non-myelinated or C fibers. Am. J. Physiol. **114:** 179, 1935.

BRONK, D. W. Fatigue in the sense organs of muscle. J. Physiol. **67:** 270, 1929.

The mechanism of the sensory end organ. Res. Publ. Assn. Res. Nerv. Ment. Dis. **15:** 60, 1934.

The activity of nerve cells. Cold Spring Harbor Symp. on Quant. Biol. **4:** 1, 1936.

BRONK, D. W. AND G. STELLA. The response to steady pressure of single end organs in the isolated carotid sinus. Am. J. Physiol. **110:** 708, 1935.

BURRILL, D. Y., F. R. GOETZEL AND A. C. IVY. The pain threshold raising effects of amphetamine. J. Dent. Res. **23:** 337, 1944.

CLARK, D., J. HUGHES AND H. S. GASSER. Afferent functions in the group of nerve fibers of slowest conduction velocity. Am. J. Physiol. **114:** 69, 1935.

FREY, M. VON. Untersuchungen ueber die Sinnesfunctionen der Menschlichen Haut. Abh. d. Math.-Physical. Classe der Konigl. Sächs. Ges. der Wiss. **23:** 175, 1896.

GALAMBOS, R. Inhibition of activity in single auditory nerve fibers by acoustic stimulation. J. Neurophysiol. **7:** 287, 1944.

GASSER, H. S. Conduction in nerves in relation to fiber types. Res. Publ. Assn. Res. Nerv. Ment. Dis. **15:** 35, 1935.

Pain producing impulses in peripheral nerves. Res. Publ. Assn. Res. Nerv. Ment. Dis. **23:** 44, 1943.

GASSER, H. S. AND J. ERLANGER. The rôle played by the sizes of the constituent fibers in a nerve trunk in determining the form of its action potential wave. Am. J. Physiol. **80:** 522, 1927.

GOLDSCHEIDER, A. Zur Lehre von den specifischen Energien der Sinnesorgane. Diss. Berlin 1881.

HARDY, J. D., H. G. WOLFF AND H. GOODELL. Studies in pain; a new method for measuring pain threshold; observations on the spatial summation of pain. J. Clin. Investigation **19:** 649, 1940.

Pain threshold in man. Res. Publ. Assn. Res. Nerv. Ment. Dis. **23:** 1, 1943.

HARTLINE, H. K. The response of single optic nerve fibers of the vertebrate eye to illumination of the retina. Am. J. Physiol. **121:** 400, 1938.

HARTLINE, H. K. AND C. H. GRAHAM. Nerve impulses from single receptors in the eye. J. Cell. Comp. Physiol. **1:** 277, 1932.

HEAD, H. AND G. HOLMES. Sensory disturbances from cerebral lesions. Brain **34:** 102, 1912.

HEINBECKER, P. AND S. H. BARTLEY. Action of ether and nembutal on the nervous system. J. Neurophysiol. **3:** 219, 1940.

HEINBECKER, P., G. H. BISHOP AND J. L. O'LEARY. Pain and touch fibers in peripheral nerves. Arch. Neurol. Psychiat. **29:** 771, 1933.

Analysis of sensation in terms of the nerve impulses. Arch. Neurol. Psychiat. **31:** 34, 1934.

HEINBECKER, P., J. L. O'LEARY AND G. H. BISHOP. Nature and source of fibers contributing to the saphenous nerve of the cat. Am. J. Physiol. **104:** 23, 1933.

HOGG, B. M. Slow impulses from the cutaneous nerves of the frog. J. Physiol. **84:** 250, 1935.

KELLGREN, J. H. On the distribution of pain arising from deep somatic structures with charts of segmental pain areas. Clin. Sci. **4:** 35, 1939.

LEWIS, T. AND E. E. POCHIN. The double pain response of the human skin to a single stimulus. Clin. Sci. **3:** 67, 1937.

LORENTE DE NÓ, R. Studies on the structure of the cerebral cortex. 11. Continuation of the study of the ammonic system. J. Psych. u. Neur. **46:** 113, 1934.

MICHELSEN, J. J. Subjective disturbances of the sense of pain from lesions of the cerebral cortex. Res. Publ. Assn. Res. Nerv. Ment. Dis. **23:** 86, 1943.

NAFE, J. P. Toward the quantification of psychology. Psych. Rev. **49:** 1, 1942.

NAFE, J. P. AND K. S. WAGONER. The nature of pressure adaptation. J. Gen. Psychol. **25:** 323, 1941.

O'LEARY, J. L. A structural analysis of the lateral geniculate nucleus of the cat. J. Comp. Neurol. **73:** 405, 1940.

PENFIELD, W. AND E. BOLDREY. Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. Brain **60:** 389, 1937.

RANSON, S. W. The course within the spinal cord of the non-myelinated fibers of the dorsal roots. J. Comp. Neurol. **23:** 259, 1913.

RIVERS, W. H. R. AND H. HEAD. A human experiment in nerve division. Brain **31:** 323, 1931.

SCHUMACHER, G. A. The influence of inflammation on the pain threshold of the skin of man. Res. Publ. Assn. Res. Nerv. Ment. Dis. **23:** 166, 1943.

SCHUMACHER, G. A. AND H. G. WOLFF. Experimental studies in headache; contrast of histamine headache with headache of migraine, etc. Arch. Neurol. Psychiat. **45:** 199, 1941.

TALAAT, M. The effect of ions on the cutaneous sensory endings of the frog. J. Physiol. **79:** 500, 1933.

TOWER, S. Units for sensory reception in the cornea. J. Neurophysiol. **3:** 486, 1940.

Pain; definition and properties of the unit for sensory reception. Res. Publ. Assn. Res. Nerv. Ment. Dis. **23:** 16, 1943.

WALKER, A. E. Central representation of pain. Res. Publ. Assn. Res. Nerv. Ment. Dis. **23:** 63, 1943.

WATTERSTON, D. Observations on sensation. J. Physiol. **77:** 251, 1933.

WEDDELL, G. The pattern of cutaneous innervation in relation to cutaneous sensibility. J. Anat. **75:** 346, 1941.

WOLFF, H. G. Some observations on pain. Harvey Lectures, 39, 1943–44.

WOLFF, H. G. AND H. GOODELL. Relation of attitude and suggestion to perception of and reaction to pain. Res. Publ. Assn. Nerv. Ment. Dis. **23:** 434, 1943.

WOLFF, H. G. AND J. D. HARDY. Studies on pain; observations on pain due to local cooling and on the factors involved in the "cold pressor" effect. Res. Publ. Assn. Res. Nerv. Ment. Dis. **23:** 123, 1943.

WOLFF, H. G., J. D. HARDY AND H. GOODELL. Measurement of effect on pain threshold of various substances. J. Clin. Investigation **20:** 63, 1941.

THE ABSORPTION OF TRIGLYCERIDE FAT FROM THE INTESTINE

A. C. FRAZER

University of Birmingham, England

The Lipolytic Hypothesis of fat absorption has been put forward in its most complete form by Verzar and McDougall (1936) (1). On the basis of a number of experimental findings which could not be reconciled with this conception of the mechanism of fat absorption, Frazer (1938) (2) suggested the Partition Hypothesis, and subsequent work has furnished added support for this alternative view. The essential points of difference between these two hypotheses are shown in table 1. It will be seen that there is general agreement on the main changes which fat may undergo during absorption. Thus, the occurrence of emulsification, hydrolysis and phosphorylation, and the importance of bile salts and the adrenal cortex, are recognised by both groups of workers. The differences lie in the way these changes are brought about, and in their significance and relationship to the absorptive mechanism. The most recent assessment of evidence (Bloor, 1943) (3) was compiled before the full publication of the experimental work upon which the Partition Hypothesis is based. Reassessment, therefore, seems desirable, and in this review the differences of opinion shown in table 1 are analysed by the critical examination of the experimental methods used, the results obtained and their interpretation in each group of problems. Attention is concentrated on the mechanism of absorption of triglyceride fats. The related problems of phospholipid and cholesterol absorption are not considered except in so far as they may be concerned with the absorption of triglyceride.

What changes does fat undergo in the intestinal lumen? The two main changes which may affect fat in the intestinal lumen are emulsification and hydrolysis. The possible extent and mechanism of each of these changes can be studied either in vitro, by experiments in which the various factors concerned are investigated under controlled, but relatively static conditions, or in vivo, by the examination of material removed from the intestine. Certain inferences, especially with regard to the functions of these changes, can only be drawn from a more extensive survey of the experimental field.

Emulsification. Ingested fat is finely emulsified in the upper part of the small intestine so that the particles average less than 0.5 μ in diameter. It is now generally accepted that the reaction of the contents of the duodenum and jejunum is acid—about pH 6.5 (4). While more alkaline material may sometimes be found, especially in the lower ileum, it is clear that the intestinal emulsifying system must be effective at pH 6.5. Adequate criteria should be applied in experiments designed to demonstrate possible intestinal emulsifying systems. The particle size and stability should be checked microscopically under dark-ground illumination, and the system should be tested under conditions similar to those prevailing in the intestine. The system should induce spontaneous emulsification without undue agitation and, if particulate absorption of fat is being

TABLE 1

Main points of difference between lipolytic and partition hypotheses

	LIPOLYTIC HYPOTHESIS (1) (3)	PARTITION HYPOTHESIS (2) (73) (47) (31) (7)
1. Changes occurring in the lumen of the intestine		
(a) Emulsification		
i. Mechanism	Acid soap	Triple complex: fatty acid/bile salt/lower glyceride (7)
ii. Function	To promote hydrolysis	To disperse unhydrolyzed fat preparatory for absorption
(b) Hydrolysis		
i. Extent	Complete	Partial.
ii. End products	Fatty acid and glycerol	Fatty acid, di- and monoglycerides and later glycerol (8)
iii. Function	Essential preliminary to absorption	Provides $\frac{2}{3}$ components of emulsifying system. Partitions fat into fatty acid and glyceride fractions
2. Passage through the outer membrane (free border) of the intestinal cell		
(a) Structure of brush border	Solid pavement	Canal structure (44)
(b) Fatty acids	Pass through the membrane as soluble complexes with bile salts (39)	Pass through the membrane either as soluble compounds or complexes
(c) Glycerides	Do not pass through the membrane	Pass through the membrane as finely dispersed emulsion of negatively charged particles less than 0.5μ in diameter (7)
(d) Adrenal cortex	Not concerned here	Controls normal electrolyte balance which is closely related to the absorption of charged particles (27)
3. Changes occurring in intestinal cell		
(a) Resynthesis	Essential part of mechanism. All fatty acid converted back to triglyceride	Not an essential part of the absorptive mechanism
(b) Phosphorylation	A stage in resynthesis of triglyceride fat (74) (75)	Not concerned with resynthesis but occurs at oil/water interface, as an essential change in interfacial structure, and probably elsewhere (10)
(c) Adrenal cortex	Essential for normal phosphorylation (55)	Not concerned in phosphorylation (61)

TABLE 1—*Concluded*

	LYPOLYTIC HYPOTHESIS (1) (3)	PARTITION HYPOTHESIS (2) (73) (47) (31) (7)
4. Distribution of absorbed fatty substances in the body		
(a) Fatty acid fraction	After resynthesis the triglyceride passes up lacteals and thoracic duct into systemic circulation Negligible amounts pass up portal vein to liver	Mainly passes by the portal vein to the liver (2) (31) (47)
(b) Glyceride fraction	Not absorbed as such. Resynthesized from fatty acid and glycerol as described above	To systemic blood via lacteals and thoracic duct and so to the fat depots (47)

considered, the droplets should be smaller than 0.5 μ in diameter, and stability should be maintained for at least 3 hours.

It has been suggested that the bile salts provide an effective emulsifying system for fat in the intestine. It can easily be shown that purified bile salts will only give crude emulsification with low stability not comparable with that known to occur in the intestine. Other authorities have attributed emulsification in the intestine to soap (5). While soap is an excellent emulsifying agent at alkaline pH, it will not cause fine emulsification at any pH level more acid than 7.5. The suggestion that acid soaps will cause fine emulsification cannot be supported experimentally and such a system is quite ineffective in the presence of calcium. Soap is a much more effective emulsifying agent if it is formed in situ at the oil/water interface. This, of course, occurs during lipolysis. Its action is also enhanced by the presence of cholesterol due to the formation of soap/cholesterol complexes (6). None of these modifications, however, will alter the pH-conditioned character of soap emulsification. If the possible factors, such as fatty acid (soap), cholesterol, bile salts and monoglycerides, are systematically examined, singly, or in double or triple combination, only one system fulfills the criteria described above and is effective over the pH range of 6.0 to 8.5. This system is the triple combination of fatty acid/bile salt/monoglyceride (7). It can also be shown that adequate quantities of monoglyceride are formed during the first 5 hours of pancreatic lipolysis both in vitro and in vivo (8).

Material from the intestinal lumen may be examined analytically, or the interfacial structure may be studied by determining the protein flocculation pattern (9) or the effect of enzymes. The emulsion is found to contain fatty acids, and tri-, di- and mono-glycerides; no phospholipid could be demonstrated. The flocculation pattern of the intestinal emulsion is that of simple negatively charged particles (10) which are not affected by Cl. welchii lecithinase (11). These findings would fit in with the conception that the fat emulsion particles in the intestinal lumen are stabilised with lower glyceride and fatty acid/bile salt complexes.

It may be concluded that of the various emulsifying mechanisms suggested, it is only the triple combination of fatty acid/bile salt/monoglyceride which has been shown to be effective under the conditions prevailing in the intestine. The various components of this system are readily available due to lipolysis and bile secretion. Phospholipid, cholesterol and other substances may, perhaps, be secondarily involved in the interfacial film, but it seems that they are not essential to the fine emulsification of fats in the intestinal lumen. The functions of this emulsification may be to provide an increased interfacial area for subsequent hydrolysis, or it may be concerned with the absorption of finely dispersed fat in the unhydrolysed state. The discussion of these two possibilities must be deferred until other points in the evidence have been reviewed.

Hydrolysis. "Lipase" may be one enzyme with many different actions, one or other of which may be emphasized by different combinations of activators and inhibitors, or it may consist of a group of esterases. For the present it is more convenient to consider these actions as being due to different enzymes, and the term "lipase" may, with advantage, be restricted to those which mainly act on the glycerides of the long-chain fatty acids. Many of the observations on the action of pancreatic lipase have been made using esters of butyric acid as the substrate. While this may be convenient for enzyme studies, it is unwise to apply such experimental findings unreservedly to the problem of long-chain triglyceride hydrolysis, since separate enzymic activity toward these various substrates can be differentiated (12), (13), (14), (15), (16) and butyric acid is water-soluble in contrast to most other fatty acids.

The main source of lipase in the intestinal lumen is the pancreas, although some may also be demonstrated in other intestinal juices or derived from bacteria (17). The initial flow of pancreatic juice, which is probably neurogenic, is rich in lipase, whereas the flow due to secretin has a poor enzyme content (18). Harper and Raper (1944) (19) have described another hormone, pancreozymin, which stimulates the production of amylase and trypsinogen; increase of lipase has also been demonstrated (20). It is generally agreed that ingested triglyceride is hydrolysed by lipase in the intestinal lumen. The extent to which hydrolysis occurs and the nature of the cleavage products in the intestine are the main points upon which opinion differs.

The percentage hydrolysis of triglyceride by pancreatic lipase in vitro is found to be influenced by the length of the experimental period, the relative amounts of oil and lipase and the degree of dispersion of the former, the presence of activators and the pH of the continuous phase, and the accumulation of cleavage products. The relationship of these factors to other evidence and their possible significance in the intestinal lumen must be assessed.

Observations of lipolysis may be made over any time period, but it is only the changes which occur during the first few hours that are of significance in relation to the absorption of fat from the intestine. It has been frequently stated that the peak of fat absorption in the human subject is reached some 6 hours after taking a meal containing fat. This is only true when excessive amounts of fat are ingested. If normal quantities, such as 30 grams are taken, the

maximum systemic lipemia usually occurs in 2 to 3 hours and the blood fat levels have returned to normal within 4½ hours (21). The delay in the case of high fat diets is due to the increase in stomach emptying time. In any case the significant time period for intestinal lipolysis is less than 6 hours, since the fat is not in the small intestine for any longer period.

The initial stages of lipolysis may be rapid, but the rate soon declines and comes almost to a stop before 30 per cent of the contained fatty acids have been liberated. The addition of more lipase causes no increase in hydrolytic rate, but addition of more substrate may do so (16), although this will not increase the percentage hydrolysis. Changing the relative proportions of oil to lipase alters the extent of hydrolysis which occurs in a given time, but the final percentage is not raised above the limiting value. The quantity of fat ingested by dogs or human subjects may be increased up to 200 to 300 grams/day without any significant diminution of the percentage absorbed (22), (23), (24). A relative increase of lipase to oil may be achieved either by the passage of small amounts of oil into the intestine, by adaptation of the composition of pancreatic juice or by the administration of additional enzyme. The increase in emptying time of the stomach due to excessive intake of fat might be an example of the former mechanism. Chylomicrographs (25), (26) made under these conditions, however, show a shift of the lipaemic curve to the right rather than any tendency for the lipaemia to be fractionated over a longer period (27). There is evidence that only the initial flow of pancreatic juice contains much lipase and that the subsequent flow consists mainly of water and salts (18). If this is the case, no great change in the relative amounts of oil and lipase will result. The evidence on the adaptation of the enzymes of the pancreatic juice to dietary intake is conflicting and this problem requires further investigation (28), (29), (30).

The addition of potent lipase to a standard fatty meal causes suppression of the normal post-absorptive lipaemia and other changes which suggest that the percentage hydrolysis has been increased (31). This can only be explained at present on the assumption that the amount of lipase normally present is very small, or that special conditions are created which allow a high percentage to occur. The limitation of hydrolysis in vitro can be readily demonstrated even if the oil is finely dispersed to a degree comparable with that found in the intestine (32). It does not seem likely therefore that fine emulsification alone can explain the occurrence of the high percentage hydrolysis required by the Lipolytic Hypothesis. Furthermore, if absorption of particulate fat occurs, the finely dispersed material will be rapidly removed from the intestinal lumen.

Willstätter and his colleagues (33) demonstrated several substances which activate lipase, of which the most important are the bile salts, albumin and calcium oleate. These substances are probably non-specific activators which help to bind the water-soluble lipase to its insoluble substrate. Elaborate systems of activators can be built up which will stabilize lipase activity. While this is necessary for the study of enzyme kinetics, the findings of such experiments are often not applicable to the systems found in the intestinal lumen. Even if these

activators are used, the percentage hydrolysis is not increased in the significant time period except under special circumstances such as those devised by Balls, Matlack and Tucker (34). There is no evidence of any complex activating mechanism necessary for lipolysis in the small intestine. Hydrolysis and absorption of fat appear to be normal in the absence of any significant quantities of albumin or calcium oleate. Absorption is interfered with in the absence of bile salts, but this can be well explained on other grounds, and the percentage hydrolysis of fat under these conditions is often normal (35).

The decreasing rate of hydrolysis has been attributed to the establishment of equilibrium due to the accumulation of cleavage products in the water phase, and it is claimed that in the intestine the removal of fatty acid and glycerol by absorption will allow hydrolysis to proceed rapidly to completion. It is doubtful, however, whether it is justifiable to apply the Law of Mass Action in this way to a non-homogeneous system where many other factors are obviously involved in the determination of hydrolytic rate. One factor which has not received sufficient attention is the relationship of the amounts of fatty acid and lipase to the available oil/water interface. As hydrolysis proceeds the interface must become loaded with oil-soluble cleavage products, especially those which contain hydrophilic groups. It is inevitable that an accumulation of fatty acid will tend to displace the enzyme from the oil/water interface with consequent interference with hydrolysis. This effect can only be avoided if the cleavage products can be removed from the interfacial film. They might be shifted either into the oil phase or into the water phase. The former could be achieved in the case of fatty acid by suppression of ionization of the carboxyl groups, but this would also result in the breaking of the oil-in-water emulsion. The fatty acid might be removed into the aqueous phase in three ways, as fatty acid, soap, or water-soluble complexes. Only the lower fatty acids, such as butyric acid, are water-soluble, and it may be significant that tributyrin behaves differently from the long-chain triglycerides on hydrolysis both in vitro and in vivo (36), (37), (38). The removal of fatty acid as soap depends on pH, and any extensive removal of fatty acid from the oil/water interface at the intestinal pH of 6.5 cannot be demonstrated in vitro (27). In a more alkaline environment this mechanism might be more effective. Fatty acids form complexes with bile salts, which are water-soluble and diffusible through artificial membranes (39). These diffusion experiments have been criticised (40), (41) but in any case the molecular relationships necessary for the formation of these complexes bear little resemblance to those prevailing in the intestinal lumen. Verzar (1) pointed out that the amount of bile salt available in the intestine was not sufficient to account for the absorption of fatty acids by this mechanism, and he suggested that this objection might be overcome if the bile salts exerted their hydrotropic action at the surface of the intestinal cell. This finds little theoretical or experimental support, and it provides no answer to the problem of the removal of fatty acid from the oil/water interface. It may be concluded that the interface will be cleared of fatty acids if they are water-soluble, or if the pH of the environment is sufficiently alkaline,

but that under other circumstances, accumulation at the interface with consequent interference with hydrolysis is likely to occur. Glycerol, if formed, will readily pass into the water phase and other possible end products such as di- and mono-glycerides are alternative substrates to triglyceride and will not cause interference.

It is generally stated that the end products of hydrolysis of triglyceride with pancreatic lipase are fatty acid and glycerol. If pancreatic lipolysis is carried out under standard conditions without the introduction of excessive quantities of activators or an abnormal pH level, it can be shown that free glycerol is not liberated during the first 5 hours. During this period fatty acid is set free and the glyceride residue shows an increase in acetyl value from 4 to 64. If this is interpreted as being due to the formation of mono- and di-glycerides, the whole of the glycerol is readily accounted for. It can also be shown in control experiments that glycerol is not destroyed during this time period, but that it does disappear over longer periods. That the end products in these experiments consist of fatty acid and mono- and di-glyceride has been further confirmed by the separation of the lower glycerides from the glyceride residue, and comparison of the compounds obtained with authentic specimens. If material from the lumen of the intestine is examined, it appears to have a comparable content of monoglyceride (8). If the end products of hydrolysis during the first 5 hours are fatty acids and glycerides, and over 90 per cent of fat is absorbed during this period, it is obvious that a significant proportion of the fat must be absorbed as unhydrolysed or partially hydrolysed triglyceride.

Material obtained from the intestinal lumen may be analysed and the proportion of fatty acid determined (42). An increase in the proportion of fatty acid has been interpreted as evidence of progressive lipolysis. Arguments have also been advanced in support of a high percentage hydrolysis on the basis of the fatty acid content of faecal fat, which is usually 70 to 80 per cent (43). In each of these instances it has been assumed that the only fatty material being absorbed is fatty acid. If this assumption is not made and the possibility of absorption of unhydrolysed fat is considered, these observations cease to have any such significance.

It is concluded that pancreatic lipase does not work in the intestine under the ideal conditions which are frequently provided in vitro. The pH of the upper part of the small intestine, the short time period and the difficulty of removal of fatty acid from the oil/water interface, will all tend to limit the extent of hydrolysis. In the lower ileum, where the reaction may be more alkaline, it is possible that more extensive hydrolysis might occur and this may act as a safeguard ensuring almost complete absorption of fat. Not only is the percentage hydrolysis likely to be restricted, but there is evidence that the individual triglycerides may be only partly hydrolysed to di- and mono-glycerides which, with the bile salts, provide an effective emulsifying system. Fine emulsification of fat will not alone influence the eventual percentage hydrolysis, but it may allow unhydrolysed fat to be removed from the intestinal lumen by particulate absorption, the evidence for which will be discussed later. If the finely dispersed fat is not removed and

the reaction of the aqueous phase becomes more alkaline or if the fat is a short-chain triglyceride, such as tributyrin, more extensive hydrolysis would be expected.

How does fat pass through the outer membrane (free-border) of the intestinal cell? The structure of the outer membrane, the so-called brush-border, of the intestinal cell has been investigated by histological methods, but interpretation under high magnification is difficult. The final conclusion on the structure of such a membrane must depend not only on its histological appearance, but also upon the results of functional permeability tests. Histological examination of the membrane has been carried out by a number of observers, whose views are well summarized by Baker (1943) (44). From the study of the intestinal cells mainly of amphibia, Baker submits convincing evidence that the brush-border contains a system of fine canals, running at right angles to the surface, through which particles rather less than 0.5 μ in diameter might pass. His preparations show these canals in both longitudinal and transverse section. The constancy of his observations, in preparations using different methods, makes it unlikely that the canals are artefacts. Wotton and Zwemer (1939) (45) have published photographs which they claim show fat passing through canals in the membrane into the intestinal cell. The physical chemical forces involved in this mechanism are obscure. It may be concluded that the possibility of a canalicular structure in the outer membrane of the intestinal cell must be borne in mind, and that it is unwise to consider problems of fat absorption on the assumption that the outer membrane has necessarily a pavement structure or that it resembles a lipoid film (46).

The passage of fat through the outer membrane of the intestinal cell is best studied by concurrent biochemical analysis of the residual material in the intestine, and histological examnation of frozen sections. The investigations should be planned on a serial basis and particular attention should be given to the length of the experimental time period and to the site from which samples are taken for section. Various types of glyceride, cleavage products, or unsaponifiable material, such as paraffin may be used. The fatty materials may be administered as a finely dispersed emulsion or as crude oil. The degree of dispersion of emulsions must be adequately checked before and during the experiment. Many emulsions must be injected intraduodenally to avoid destruction by the acid in the stomach.

Fat must pass through the membrane either as glyceride or as fatty acid. It is agreed that lipolysis occurs in the intestinal lumen, and the fatty acid is likely to be absorbed, provided it can be separated from the residual unhydrolysed glyceride as a water soluble compound or complex. The precise mechanism by which fatty acid passes through the membrane is not known, but it is probably the same as for other water-soluble compounds. The supporters of the lipolytic hypothesis claim that all the fat absorbed from the intestine must be hydrolysed to fatty acid and glycerol before it can pass into the intestinal cell. It has already been shown that hydrolysis may be restricted in the intestinal lumen and that triglycerides may be only partially broken down. Further, differentiation be-

tween the absorption of triglyceride and its constituent fatty acids can be demonstrated (47) and will be discussed in more detail later. For these reasons the possibility of absorption of unhydrolysed fat must be considered.

Glycerides might be absorbed either in molecular dispersion or in particulate form. Glycerides might be molecularly dispersed in water in the form of phosphorylated compounds. The available evidence suggests that phospholipids are broken down in the intestinal lumen, and that synthesis of these compounds occurs after the passage of the fatty components into the intestinal cells. No hydrotropes have been demonstrated which form an association with glycerides, so that molecular dispersion in complex form does not seem possible. It must be concluded on present evidence that glycerides do not pass through the intestinal membrane in a state of molecular dispersion.

Although it has been claimed that large globules of fat pass through the intestinal membrane (45), it is difficult to accept this mechanism on histological evidence alone in the absence of any adequate physical chemical explanation. If the possibility of finely dispersed fat particles passing through the membrane is to be considered, it must be demonstrated that an effective emulsifying system exists in the intestinal lumen, that the membrane structure is suitable, and that finely dispersed unsaponifiable material can be absorbed. The physical chemical forces required for the passage of the particles into the cell must be defined, and the relationship of this mechanism to the selective absorption of other foodstuffs determined.

The existence of an effective emulsifying system in the intestine, which will reduce the particle diameter to less than 0.5 μ, has already been discussed. The canalicular structure of the outer membrane of the intestinal cell described by Baker would seem to be suitable for absorption of particles of this size. The absorption of unsaponifiable material, such as paraffin, has been the subject of much controversy. Channon and his co-workers (48), (49) have demonstrated the absorption of paraffins, but other workers have denied that absorption occurs (50), (51), (52). Unfortunately these observations were made without adequate control of the state of division of the paraffin. Statements on the absorption of emulsified material (53) are often misleading, since the milky appearance of oil-in-water emulsions, especially if they are made with pre-formed soap solution, cannot be regarded as a satisfactory criterion of fine emulsification. The size of the particles in these experiments must be determined with reasonable accuracy by microscopic methods. It is generally agreed that absorption of paraffin is minimal if it is administered orally either unemulsified or as a coarse emulsion. If a paraffin emulsion with an average particle diameter of less than 0.5 μ is injected intraduodenally, it is absorbed in amounts comparable with a similar olive oil emulsion. The absorption of these paraffin emulsions has been demonstrated by residual analyses and histological studies, and has been a constant finding in large groups of rats (7). It has been concluded in the past that paraffin was not absorbed because it was not hydrolysed. Although the reasons for reaching this conclusion were probably false, it is possible that the conclusion tself was correct, for hydrolysis provides two essential components of the

intestinal emulsifying system. Paraffin is thus only absorbed from the intestine if it is finely dispersed with suitable emulsifying agents before administration, and dispersion is maintained in the alimentary tract.

The fat particles in the intestinal lumen can be shown to carry a negative charge. Teorell (54) demonstrated the relationship of electrolyte diffusion to the passage of negatively charged particles through membranes. It is not impossible that the outer membrane of the intestinal cell may be charged due to difference in the diffusion rates of absorbed ions and that this may have an important bearing on the absorption of finely dispersed fat particles. The relationship of such a system to the selective absorption of other substances than fat requires investigation.

If electrolyte absorption is closely associated with the passage of fat particles through the outer membrane, it might be expected that double adrenalectomy, with the consequent electrolyte disturbance, might affect fat absorption. Verzar and Laszt (1935) (55) claimed that fat absorption in rats was reduced to less than 10 per cent of normal following double adrenalectomy. They considered that this depression of fat absorption was due to interference with phosphorylation. A number of workers have confirmed that fat absorption is depressed after adrenalectomy. Normal absorption was re-established by the administration of corticosterone and it was also improved by adequate sodium therapy (56), (57), (58), (59). These and other observations (60), (61) do not substantiate the view that the adrenal cortex is concerned with phosphorylation processes in the intestinal cell. In a study of the differential absorption of triglycerides and fatty acid in adrenalectomised rats (27) we found that triglyceride absorption was depressed to 40 per cent, but that fatty acid absorption was practically normal quantitatively. We confirmed Verzar's observation that fatty acid absorption in adrenalectomised animals gave rise to toxic symptoms. The absorption of triglyceride was improved in salt-treated animals. The time of onset of defective fat absorption coincided with the development of electrolyte imbalance. Bavetta and Deuel (1942) (62) have described normal absorption of tributyrin and butyric esters after adrenalectomy, and they suggest that the adrenal glands play a rôle in the absorption of the long-chain fatty acids. It has already been suggested that tributyrin is likely to be absorbed more as fatty acid than unhydrolysed fat, whereas long-chain glycerides will tend to pass through mainly in particulate form. If adrenalectomy interferes with the absorption of fat particles but not fatty acid, the observations of Bavetta and Deuel are readily explained.

It may be concluded from these various observations that fat can be absorbed as fatty acid provided that the cleavage products can be removed into the aqueous phase from the oil/water interface. This appears to be more easily accomplished in the case of butyric acid than the long-chain fatty acids. It is suggested that the possibility of particulate absorption of unhydrolysed fat must be considered, especially in relation to the absorption of long-chain fats. The absorption of fat particles may be dependent upon normal electrolyte metabolism, and this might account for the importance of the adrenal cortex in the absorption of long-chain compounds.

These two mechanisms of absorption can be accepted on the evidence available without necessarily assuming that they both play a part in normal fat absorption, although it seems probable that this is the case. The relative proportions of fat normally absorbed in the hydrolysed or unhydrolysed state have not yet been determined, but it seems likely that the ratio will vary according to the type of fat ingested, the conditions in the intestine, the normal functioning of the intestinal cells and other metabolic activities. The effect of added lipase (31), (47) and the differentiation of triglyceride and fatty acid absorption suggests that, under normal circumstances, most of the fat in the chyle is derived from material absorbed in particulate form. Munk and Rosenstein (63), in their classical experiments, recovered only 60 per cent of ingested fat from a lymphatic fistula; in adrenalectomised animals interference with particulate absorption causes a depression to 40 per cent of normal, and lipolysis in the significant time period tends to be restricted to about 30 per cent under the conditions found in the intestine. Bearing in mind that an upset of the normal mechanism of particulate absorption may lead to an increase in the percentage hydrolysis, it seems likely that, so far as long-chain triglycerides are concerned, at least 60 per cent of the fat absorbed may pass through the outer membrane of the intestinal cell as finely dispersed particles of unhydrolysed fat.

What changes does fat undergo in the intestinal cell? Evidence of changes in the composition of fatty material in the intestinal cell may be obtained directly by histo-chemical studies or indirectly by comparison of the substances entering and leaving the cell. This evidence may be correlated with in vitro experiments which may indicate both the nature and possible function of the changes observed. The interpretation of histo-chemical methods is difficult, not only for technical reasons, but also because they give a static picture of intracellular changes. The available methods for the differentiation of triglycerides from fatty acids are unreliable (64). The examination of substances immediately before entering or after leaving the intestinal cell is not possible, but an approximate assessment is made by the study of the contents of the intestinal lumen and the blood or lymph leaving the intestine. It is obviously an advantage if the material is collected as soon as possible after leaving the intestinal cell to minimize the danger of admixture with entirely irrelevant substances. Thus, the technique of collecting intestinal lymph (65) is preferable to thoracic duct canulation. Whatever methods are used, however, considerable caution must be exercised in their interpretation.

The two main changes which are thought to occur in the intestinal cell are resynthesis of triglyceride and phosphorylation. There is no doubt that most of the fat in the chyle is triglyceride. If complete hydrolysis of fat occurs before absorption, then resynthesis must take place. If, however, a large proportion of the fat is absorbed in the unhydrolysed state, then resynthesis is not necessarily important. Various experiments have been made which purport to show triglycerides in the chyle derived from monoglycerides, fatty acids or other esters (66), (67), (68). In none of these experiments can the synthesis of triglycerides in the lumen of the intestine or the possibility of the triglycerides being derived

from some other source than immediately absorbed material, be excluded. That glyceride synthesis from fatty acids and glycerol may occur in the presence of lipase can be readily demonstrated in vitro (69), (70), (71), and it seems reasonable to suppose that given suitable conditions similar changes can occur in vivo. Whatever may be the true interpretation of the experiments so far devised to demonstrate resynthesis, they do no more than indicate that synthesis can occur under certain conditions. They do not show the extent of triglyceride resynthesis during normal fat absorption. Verzar has advanced the histological evidence of Jeker (72) as support for resynthesis. He demonstrated what appeared to be fatty acid in the intestinal cell 30 minutes after ingestion, but after 6 hours the cells were seen to be full of neutral fat. Quite apart from the validity of the histo-chemical methods employed, these experiments cannot be accepted as evidence of resynthesis, for identical results could obviously be obtained by early fatty acid absorption and later absorption of unhydrolysed fat (73).

Phosphorylation has been regarded as an important intermediate step in triglyceride resynthesis (74), (75), (1). Whether resynthesis occurs or not, the conception that phosphorylation is a stage in this process is open to question. The reversible nature of lipolysis has been amply demonstrated in vitro, and the formation of triglyceride from cleavage products in the presence of lipase does not appear to require intermediate phosphorylation. The available evidence suggests that phospholipid in the intestinal cell should be more properly regarded as an end product than an intermediate compound in resynthesis.

Although the relationship between phosphorylation and triglyceride resynthesis is doubted, there is good evidence that phospholipids are formed in the intestinal cells. This conclusion is based on histo-chemical studies (76) and on experiments in which labelled phosphorus or fatty acids were traced from the intestinal lumen through the intestinal cell into phospholipids leaving the intestine (77), (78), (79), (80), (81).

Verzar supported this conception of the importance of phosphorylation by experiments which showed that moniodo-acetic acid and phloridzin (82) caused interference with absorption of fat. The doses used were excessively large, and can be shown to cause extensive destruction of the intestinal mucosa (83). It seems unreasonable to assume, therefore, that the interference with fat absorption in these experiments can be attributed to faulty phosphorylation. Even if phospholipids are not concerned with resynthesis, there may be many other important reasons for phospholipid formation. Phospholipid is a convenient transportable form of fat and it may be more suitable than triglyceride for many metabolic reactions. It can also be demonstrated that phospholipid is an essential component of the stabilising film of fat particles circulating in the blood stream, and that the maintenance of the phospholipid content of the interfacial film is necessary to their continued dispersion. Negatively charged fat particles flocculate in the presence of plasma proteins and 1 per cent sodium chloride at a pH of 7.5 (9). This flocculation does not occur if the particles are stabilised with phospholipid instead of soap. If these two types of emulsion are treated with Cl. welchii lecithinase, the soap-stabilised emulsion is unaffected while the phos-

pholipid-stabilised emulsion flocculates. If the fat particles in the intestinal lumen are examined they are found to behave as simple negatively-charged particles, but the chylomicrons on the other hand behave similarly to a phospholipid stabilised emulsion. It is tentatively suggested that phospholipid is added to the fat particles in the intestinal cell, and it may be formed in situ at the oil/water interface. It is not impossible that monoglyceride might provide the basis for this phospholipid formation (10). Alternatively, phospholipid may be preformed in the intestinal cell and spread as a film on the surface of the oil globule.

The relationship of the adrenal cortical hormones to phosphorylation (55) has not been substantiated by other workers either in connection with fat (61), carbohydrate (84), (85), or vitamin (86) absorption. It is difficult, therefore, to accept this conception of the action of the adrenal gland, and a possible alternative explanation of its relationship to fat absorption, due to the associated electrolyte disturbance, has already been discussed. It may be concluded that the question of the significance of resynthesis in the intestinal cell depends upon the form in which fat is absorbed. If a large proportion of fat enters the cell as triglyceride, resynthesis ceases to be of such vital importance. The relationship of phosphorylation to resynthesis or adrenal cortical hormone is open to question. It may, however, represent an essential change in the interfacial structure of fat particles needed to maintain normal stability and dispersion in the blood stream. It is likely that phospholipid formation has other important functions not yet clearly defined.

What is the distribution of absorbed fatty material? This problem can only be studied by in vivo experiments. The fatty material may be labelled by staining, by the introduction of recognisable chemical groups, or the incorporation of radio-active elements. In addition to determining the immediate destination of the fatty material, the pathways along which it passes from the intestine may be studied. The two main routes are the lacteal-lymphatic channel through which the chyle is passed into the systemic circulation, and the capillary network which leads via the portal vein to the liver. While fat, travelling in the systemic blood, need not necessarily be removed into the fat depots, or that in the portal blood into the liver, it seems probable that systemic lipaemia will tend to be associated with increased deposition in the depots and portal lipaemia with a rise in liver fat, provided that excessive quantities of fatty materials are not employed. In the study of the pathways of absorption repeated chemical analyses may be used, but the number of observations made is frequently restricted. Chylomicrographs, for which much less blood is required, give a comparative serial picture of blood fat changes which is often more valuable than occasional chemical analysis in the study of fat absorption (25), (21), (26), (87).

According to the Lipolytic Hypothesis fat is completely hydrolysed before absorption and all the fatty acid is reconverted into triglyceride in the intestinal cell and passes by the lacteal lymphatic route into the systemic blood. Under these circumstances it would be expected that a reasonable similarity would exist between the phenomena observed during absorption after the ingestion of triglyceride or the administration of equivalent amounts of its constituent fatty acids

and glycerol. This was not so, and constant differences could be demonstrated between groups of animals fed on olive oil and those receiving oleic acid and glycerol. In the former, the intestinal cells were filled with fat particles, the lacteals had a milky appearance, there was the characteristic post-absorptive systemic lipaemia, and Sudanised oil could be traced to the fat depots; the portal blood fat was not appreciably changed and only slight deposition of Sudanised material could be seen in the liver. In the animals receiving fatty acid, the intestinal cells had a granular appearance, there was no milkiness of the lacteals, there was no post-absorptive systemic lipaemia and no deposition of Sudanised material in the fat depots; the portal blood, however, showed an increase of fatty material and extensive deposition of Sudanised fat could be seen in the liver (47). In accord with these observations it was also shown that the addition of extra lipase to triglyceride accentuated the fatty acid component of the absorption picture, and that inhibition of lipolysis with sodium hexadecyl sulphate produced opposite effects. In the human subject the addition of lipase caused a marked reduction in the post-absorptive systemic lipaemia (31).

Chemical analysis of the portal blood is of limited value in determining the use of this pathway, since small changes may be significant and the relationship to the type of fat administered, the degree of hydrolysis and the time of sampling, may profoundly influence the interpretation of the results. In spite of these difficulties, however, evidence has been put forward which indicates that fat may be absorbed via the portal vein. Numerous observers (88), (89), (90), (91), (92) have demonstrated an increase of fat in the portal blood during absorption, but contrary evidence is presented by Zucker (1920) (93), and by other workers (94), (95). Raper (1913) (96) suggested that the lower fatty acids might be absorbed by a different mechanism from long-chain compounds, which is of particular interest in the light of more recent work (36), (37), (38), (62) and the points previously discussed in this paper, which indicate the probability that the short chain fatty acids will be more easily detached from the oil/water interface. These various observations, and the apparent disappearance of 40 per cent of fat in Munk's experiment, the demonstration of 70 per cent absorption of fat in some cases of idiopathic steatorrhoea without any appreciable systemic lipaemia, the fact that tributyrin is absorbed but cannot be demonstrated in the chyle (37) or in the fat depots (36) all indicate the presence of an alternative pathway which is likely to be the portal vein. It is concluded, therefore, that fat may pass by the portal vein and that this pathway is used when fatty acid is absorbed. During triglyceride absorption, it is the short-chain fatty acids which are most likely to be concerned in this mechanism. The observations of Winter and Crandall (94) have been interpreted as being contrary to the Partition Hypothesis. They showed that there was no significant increase of fatty material in the portal blood of dogs during the absorption of olive oil from the intestine. These findings are completely in accord with the Partition Hypothesis and with the observations made in rats in which minimal changes were found in the portal blood and liver after administration of olive oil in contrast to the marked increase of fatty material which occurred after feeding oleic acid (47).

Extensive information on the composition of the fat depots is now available (97). It is clear that depot fat may be derived partly from food fat and partly from other sources. The resemblance between the fat in the depots and the food, especially on a high fat diet, may be explained by either hypothesis of fat absorption. The apparent selection of certain fatty acids for deposition in the depots is a complex problem. It is possible that further studies of lipolysis, the behaviour of different fatty acids at the oil/water interface and the factors concerned in fat synthesis, may throw some light on this mechanism. The occurrence of dehydrogenases in the intestinal lumen has been the subject of conflicting reports (98), (99), (100), (101), (102), (103), and for this reason has been excluded from this discussion. It is clear, however, that this might profoundly influence the hydrolysis, absorption and distribution of fatty material.

From the evidence submitted in this review, it is suggested that the Lipolytic Hypothesis of fat absorption, in its present form, fails to explain an increasing number of observations in this field, and many of the assumptions upon which it is based are in need of reconsideration. In particular, this conception of the fat absorption mechanism maintains that fat is completely hydrolysed in the intestinal lumen, that the intestinal cell has an outer pavement membrane, that paraffins are not absorbed, that the adrenal glands control phosphorylation of fat in the intestinal cell and that no significant amount of fatty material passes up the portal vein during absorption. None of these points is satisfactorily supported by the available evidence. In addition, the Lipolytic Hypothesis provides no adequate explanation of the fine emulsification of fat in the intestinal lumen, the differentiation between neutral fat and fatty acid absorption, the effect of added or inhibited lipolysis, the difference between the absorption of tributyrin and long-chain triglycerides, and the improvement in fat absorption when adrenalectomised animals are adequately salt-treated. The Partition Theory has been put forward as an alternative working hypothesis upon which further investigation of the many outstanding problems of fat absorption might be based.

REFERENCES

(1) Verzar, F. and E. J. McDougall. Absorption from the intestine. Longmans, Green & Co., London, 1936.
(2) Frazer, A. C. Analyst **63**: 308, 1938.
(3) Bloor, W. R. Biochemistry of fatty acids. Reinhold Pub. Corp., New York, 1943.
(4) Robinson, C. S. J. Biol. Chem. **108**: 403, 1935.
(5) Mellanby, J. J. Physiol. **64**: 5P, 1927.
(6) Schulman, J. H. and E. G. Cockbain. Trans. Faraday Soc. **36**: 651, 1940.
(7) Frazer, A. C., J. H. Schulman and H. C. Stewart. J. Physiol. **103**: 306, 1944.
(8) Frazer, A. C. and H. G. Sammons. Biochem. J. **39**: 122, 1945.
(9) Elkes, J. J., A. C. Frazer, J. H. Schulman and H. C. Stewart. Proc. Roy. Soc. A. **184**: 102, 1945.
(10) Elkes, J. J. and A. C. Frazer. J. Physiol. **102**: 24P, 1943.
(11) Macfarlane, M. G. and B. C. J. G. Knight. Biochem. J. **35**: 885, 1941.
(12) Willstätter, R. and F. Memmen. Hoppe-Seyler's Ztschr. **138**: 216, 1924.
(13) Loevenhart, A. S. J. Biol. Chem. **2**: 397, 415, 427, 1907.
(14) Cherry, I. S. and L. A. Crandall. Am. J. Physiol. **100**: 265, 1932.

(15) Wolvekamp, H. P. and K. Z. Griffioen. Ztschr. phys. Chem. **223**: 36, 1934.
(16) Sobotka, H. and D. Glick. J. Biol. Chem. **105**: 199, 1934.
(17) Michaelis, L. and Y. Nakahara. Ztschr. Immun. Forsch. **36**: 449, 1923.
(18) Mellanby, J. J. Physiol. **60**: 85, 1925.
(19) Harper, A. A. and H. S. Raper. J. Physiol. **102**: 115, 1943.
(20) Greengard, H., M. I. Grossman, J. R. Woolley and A. C. Ivy. Science **99**: 350, 1944.
(21) Frazer, A. C. and H. C. Stewart. J. Physiol. **90**: 18, 1937.
(22) Rubner, M. Ztschr. Biol. **15**: 115, 1879.
(23) Arnschink, L. Ztschr. Biol. 434, 1890.
(24) Rekers, P. E., J. C. Abels and C. P. Rhoads. J. Clin. Investigation **22**: 243, 1943.
(25) Fish, P. A. and S. Gage. Am. J. Anat. **34**: 1, 1924.
(26) Frazer, A. C. and H. C. Stewart. J. Physiol. **95**: 21P, 23P, 1939.
(27) Frazer, A. C. Unpublished.
(28) Bompiani, R. Arch. ital. Biol. **58**: 111, 1913.
(29) Comfort, M. W. and A. E. Osterberg. Am. J. Digest Dis. **8**: 337, 1941.
(30) Grossman, M. I., H. Greengard and A. C. Ivy. Am. J. Physiol. **138**: 676, 1943; **141**: 38, 1944.
(31) Frazer, A. C. J. Physiol. **102**: 329, 1943.
(32) Frazer, A. C. and V. G. Walsh. J. Physiol. **78**: 467, 1933.
(33) Willstätter, R., E. Waldschmidt-Leitz and F. Memmen. Hoppe-Seyler's Ztschr. **125**: 43, 1922.
(34) Balls, A. K., M. B. Matlack and I. W. Tucker. J. Biol. Chem. **122**: 125, 1937.
(35) Hutchinson, H. S. and G. B. Fleming. Glasgow Med. J. **94**: 65, 1920.
(36) Eckstein, H. C. J. Biol. Chem. **81**: 613, 1929.
(37) Hughes, R. H. and E. J. Wimmer. J. Biol. Chem. **108**: 141, 1935.
(38) Davis, R. E. J. Biol. Chem. **88**: 67, 1930.
(39) Verzar, F. and A. Kuthy. Biochem. Ztschr. **205**: 369; **210**: 265, 281, 1929; **230**: 451, 1931.
(40) Breusch, F. L. Biochem. Ztschr. **293**: 880, 1937.
(41) Vonk, H. F., C. Engel and C. Engel. Biochem. Ztschr. **295**: 171, 1938.
(42) Levites, S. Hoppe-Seyler's Ztschr. **49**: 273, 1906.
(43) Fowweather, F. S. Brit. J. exper. Path. **7**: 15, 1926.
(44) Baker, J. R. Quart. J. micr. Sci. **84**: 73, 1942.
(45) Wotton, R. M. and R. L. Zwemer. Anat. Rec. **75**: 493, 1939.
(46) Davson, H. and J. F. Danielli. The permeability of natural membranes. The University Press, Cambridge.
(47) Frazer, A. C. J. Physiol. **102**: 306, 1943.
(48) Channon, H. J. and G. A. Collinson. Biochem. J. **22**: 391, 1928; **23**: 676, 1929.
(49) Channon, H. J. and J. Devine. Biochem. J. **28**: 467, 1934.
(50) Henriques, V. and C. Hansen. Zentralbl. Physiol. **14**: 313, 1900.
(51) Mellanby, J. J. Physiol. **64**: 33P, 1927.
(52) Bloor, W. R. J. Biol. Chem. **15**: 105, 1913.
(53) Verzar, F. Ergebn. Physiol. **32**: 447, 1931.
(54) Teorell, T. Trans. Faraday Soc. **33**: 940, 1937.
(55) Verzar, F. and L. Laszt. Biochem. Ztschr. **276**: 11, 396, 1935.
(56) Barnes, R. H., E. S. Miller and O. Burr. J. Biol. Chem. **140**: 241, 1941.
(57) Clark, W. G. and A. N. Wick. Proc. Soc. exper. Biol. N. Y. **42**: 336, 1939.
(58) Bavetta, L. A., L. Hallman, H. J. Deuel, Jr. and P. O. Greeley. Am. J. Physiol. **134**: 619, 1941.
(59) Clark, W. G. and E. M. MacKay. Am. J. Physiol. **137**: 104, 1942.
(60) Ochoa, S. and R. J. Rossiter. J. Physiol. **97**: 1P, 1940.
(61) Stillman, N., C. Entenman, E. Anderson and I. L. Chaikoff. Endocrinology **31**: 481, 1942.

(62) Bavetta, L. A. and H. J. Deuel, Jr. Am. J. Physiol. **136:** 712, 1942.
(63) Munk, I. and A. Rosenstein. Virchow's Arch. **123:** 230, 484, 1891.
(64) Kaufmann, C. and E. Lehman. Virchow's Arch. **270:** 360, 1928.
(65) Wilbrandt, W. and H. Süllmann. Biochem. Ztschr. **270:** 63, 1934.
(66) Franck, O. Ztschr. Biol. **36:** 568, 1898.
(67) Argyris, A. and O. Franck. Ztschr. Biol. **59:** 143, 1912.
(68) Müller, J. and H. Murschhauser. Biochem. Ztschr. **78:** 63, 1917.
(69) Pottevin, H. C. R. Acad. Sci. Paris **136:** 767, 1903.
(70) Artom, C. and L. Reale. Bull. Soc. Chim. biol. Paris **18:** 959, 1936.
(71) Armstrong, H. E. and H. W. Gosney. Proc. Roy. Soc. B **88:** 176, 1914.
(72) Jeker, L. Pflüger's Arch. **237:** 1, 1936.
(73) Frazer, A. C. Physiol. Rev. **20:** 561, 1940.
(74) Sinclair, R. G. J. Biol. Chem. **82:** 117, 1929.
(75) Sinclair, R. G. Oil and Soap **15:** 70, 1938.
(76) Lison, L. Histochemie Animale. Gauthier-Villars, Paris, 1936.
(77) Sinclair, R. G. J. Biol. Chem. **115:** 211, 1936.
(78) Artom, C. and G. Perretti. Arch. int. Physiol. **42:** 61, 1935.
(79) Perlman, I., S. Ruben and I. L. Chaikoff. J. Biol. Chem. **122:** 169, 1938.
(80) Hahn, L. and G. Hevesy. C. R. Lab. Carlsberg **22:** 188, 1938.
(81) Hevesy, G. Enzymologia **5:** 138, 1938.
(82) Verzar, F. and L. Laszt. Biochem. Ztschr. **270:** 35, 1934.
(83) Wood, H. O. Proc. Soc. exper. Biol., April, 1944.
(84) Deuel, H. J., Jr., L. F. Hallman, S. Murray and L. T. Samuels. J. Biol. Chem. **119:** 607, 1937.
(85) Althausen, T. L., E. M. Anderson and M. Stockholm. Proc. Soc. exper. Biol. N. Y. **40:** 342, 1939.
(86) Clark, W. G. Endocrinology **28:** 545, 1941.
(87) Elkes, J. J., A. C. Frazer and H. C. Stewart. J. Physiol. **95:** 68, 1939.
(88) Errico, G. Arch. Fisiol. **4:** 513, 1907.
(89) Roger, H. and L. Binet. C. R. Soc. Biol. Paris **86:** 79, 1922.
(90) Sicard, J. A., R. Fabre and G. Forestier. Bull. Soc. Chim. biol. Paris **5:** 413 1923.
(91) Cantoni, O. Boll. Soc. ital. Biol. sper. **3:** 1278, 1928.
(92) Nedwedski, S. Pflüger's Arch. **214:** 337, 1926.
(93) Zucker, T. F. Proc. Soc. exper. Biol. N. Y. **17:** 89, 1920.
(94) Winter, I. C. and L. A. Crandall. J. Biol. Chem. **140:** 97, 1941.
(95) Little, J. M. and C. S. Robinson. Am. J. Physiol. **134:** 773, 1941.
(96) Raper, H. S. J. Biol. Chem. **14:** 117, 1913.
(97) Hilditch, T. P. The chemistry of natural fats. Chapman & Hall Ltd., London, 1940.
(98) Long, J. H. and F. Fenger. J. Am. Chem. Soc. **39:** 1278, 1917.
(99) Tangl, H. and N. Berend. Biochem. Ztschr. **220:** 234, 1930.
(100) Berend, N. Biochem. Ztschr. **260:** 490, 1933.
(101) Page, H., V. Pasternack and L. Burt. Biochem. Ztschr. **223:** 445, 1930.
(102) Bodansky, M. Proc. Soc. exper. Biol. N. Y. **28:** 630, 1931.

THE USE OF ISOTOPICALLY MARKED CARBON IN THE STUDY OF INTERMEDIARY METABOLISM

JOHN M. BUCHANAN AND A. BAIRD HASTINGS

The Department of Physiological Chemistry, University of Pennsylvania, and The Department of Biological Chemistry, Harvard University

The contribution of physics to the analytical armamentarium of chemistry in the form of isotopes has greatly extended the vision of biochemists into the understanding of the metabolism of both organic and inorganic substances in biological systems. The present paper will review those experiments on the metabolism of organic compounds and of CO_2 in which isotopes have been employed to trace the biological adventures of carbon atoms.

Two isotopes of carbon have been available for these studies, the stable isotope, C^{13}, often called heavy carbon, and the radioactive isotope, C^{11}. In many instances, carbon atoms have been tagged with stably bound deuterium and these isotopically marked organic compounds have been successfully used in the study of biological reactions.

Although the use of these three types of isotopic compounds has resulted in many new and interesting facts of metabolism, they have served an even greater function by enabling biochemists to integrate and reinterpret information obtained by conventional analytical methods. It is, therefore, the intent of this review to show how carbon isotopes have been of aid in developing certain branches of knowledge rather than to divide the subject matter into experiments carried out with the several isotopes. In this way, it is believed that a clearer appreciation can be gained of what has been accomplished as well as what problems need next be attacked.

No attempt has been made to cover thoroughly the field of bacterial metabolism since it has previously been reviewed by Werkman and Wood (145). Certain recent publications on this subject along with other older material which is pertinent to the discussion of mammalian metabolism will be included, however.

The rôle that isotopic carbon has played in elucidating certain reactions of mammalian metabolism will be discussed in the light of Schoenheimer's theory of the dynamic state of the body constituents (115). Since carbon isotopes have been used particularly in the study of carbohydrate and fat reactions, special emphasis will be placed on the metabolism of these compounds.

METABOLISM IN THE LOWER FORMS. The detailed investigation with isotopic carbon of the utilization of CO_2 by bacterial and mammalian cells has removed one of the oldest distinctions between plant and animal life. Although the extent of CO_2 assimilation and subsequent history of the carbon atoms in these two forms of life vary greatly, the occurrence of this remnant of a photosynthetic process in respiratory reactions has required the adoption of an altered point of view toward the significance of CO_2 in metabolism.

A. *Metabolism of Carbon Dioxide in Plants.* The use of isotopic radioactive

carbon (C^{11}) as a tool of biochemical research was first introduced by Ruben, Hassid and Kamen (110). In the course of their studies on photosynthesis, Ruben and co-workers (113) isolated from plants (*Chlorella pyrenoidosa*) exposed to an atmosphere of radioactive CO_2, a radioactive fraction containing carboxyl and alcoholic hydroxyl groups. Plants which were allowed to grow in the presence of light produced a photosynthetic intermediate which contained only a small fraction of the total C^{11} assimilated in the carboxyl groups; most of the activity resided in non-carboxyl carbon atoms of the molecule. In contrast, when plants were exposed to $C^{11}O_2$ in the absence of light, a fraction was isolated which contained most of the radioactivity in the carboxyl groups. This finding indicates that CO_2 fixation occurs during the dark reaction but that the light reaction is responsible for the reduction of the carboxyl groups formed during the dark reaction. It should be noted, however, that although the use of isotopic carbon provides the most direct evidence for CO_2 fixation in the dark reaction, previous experiments by workers in the field of photosynthesis had led them to a similar conclusion. This is thoroughly reviewed by Franck and Gaffron (54).

Data obtained by ultracentrifuge and diffusion observations (112, 114) on the C^{11}-containing intermediate have demonstrated that this substance is an ionized molecule with a molecular weight of approximately 1,000 to 1,500. During the period of photosynthesis, there is an increase in molecular weight. Despite the high sensitivity of the radioactive indicator method, the first molecules in which radioactivity was detected were of high molecular weight. No trace of the presence of molecules of small molecular weight such as formic acid could be detected.

Some of the results obtained by the Berkeley workers are not readily interpretable (112) on the basis of the classical theories of Willstäter and Stöll which are founded upon two main assumptions—namely, 1, the formation of an addition compound between CO_2 and chlorophyll, and 2, the subsequent photochemical reduction to formaldehyde.

From studies using C^{11}, Ruben, Kamen and co-workers support a theory of photosynthesis involving the following steps:

1. The first step in photosynthesis is the addition of CO_2 to a large molecule of unknown composition with the formation of a carboxyl group. This reaction is a thermal (non-photochemical) process which is probably enzymically catalyzed. The uptake of carbon dioxide in this dark reaction is initially rapid and gradually decreases, suggesting that the reaction may be reversible. Chlorophyll is considered not to be concerned with this reaction, since no relationship was found between the concentration of chlorophyll in the leaves and the amount of CO_2 assimilation during the dark reaction. This first reaction is represented by the following equation:

$$RH + CO_2 \rightleftharpoons RCOOH \qquad (1)$$

2. The second, irreversible, endothermic reaction of photosynthesis is considered to be catalyzed by chlorophyll. It is represented by the equation:

$$RCOOH + H_2O \xrightarrow{\text{chlorophyll}} RCH_2OH + O_2 \qquad (2)$$

Although the work of Ruben and Kamen has demonstrated that a reversible carboxylation reaction is an initial step of photosynthesis, it has been difficult to formulate the details of the reactions involved. The nature of the substance "R" with which CO_2 combines is unknown. Ruben and Kamen have indirect evidence, however, suggesting that RH may be an aldehyde. A more serious obstacle, however, has been the apparent irreversibility of known decarboxylation reactions, the reversibility of which could provide a mechanism for the lengthening of a carbon chain ultimately to a hexose unit. This difficulty has been partially overcome by the demonstration by Lipmann and Tuttle (83) and by Utter, Lipmann and Werkman (135) that the phosphoroclastic reaction may be reversed:

$$CH_3COCOOH + H_3PO_4 \leftrightharpoons CH_3COOPO_3H_2 + HCOOH \qquad (3)$$

On the basis of these investigations, Lipmann and Tuttle (83) have proposed a theory similar in many respects to an earlier one of Ruben (109), whereby carbon chains are progressively lengthened by a series of alternating phosphorylation and photoreduction reactions:

$$R\cdot COOH + \text{ad.} \sim PO_3^{-1} \rightarrow R\cdot CO\cdot OPO_3^{=} + \text{ad.H.} \qquad (4)$$

$$R\cdot COOPO_3^{=} + CO_2 + H_2 \| O \rightarrow R\cdot CO\cdot COOH + HO\cdot PO_3^{=} + \frac{O_2}{2} \qquad (5)$$

$$R\cdot CO\cdot COOH + H_2 \| O \rightarrow R\cdot CHOH\cdot COOH + \frac{O_2}{2} \qquad (6)$$

Further phosphorylation of the carboxyl group of the compound R·CHOH·COOH, reductive carboxylation and hydrogenation to hydroxy acid would result in the formation of the next higher homologue, R·CHOH·CHOH·COOH.

The reducing reagent in these reactions is photolytically formed hydrogen. In addition to its reductive functions, the reoxidation of one out of seven moles of photolytic hydrogen is presumed to be the source of energy required to supply the essential energy-rich phosphate bonds.

For more extensive information on photosynthetic reactions and the rôle which isotopic compounds have played, the reader is referred to reviews by Franck and Gaffron (54, 55), and by van Niel (97, 98).

In an extension of their work with plants, Ruben and Kamen (111) found that fresh yeast cells suspended in water also assimilate CO_2. In their experiments, approximately one CO_2 molecule was reduced for every fifty molecules produced in respiration. Though most of the radioactivity was present in the precipitate formed by the addition of barium ions, the activity did not reside in the carboxyl groups of the barium precipitable material.

B. *Metabolism of Carbon Dioxide in Microorganisms.* Although certain forms of autotrophic bacteria may utilize CO_2 as the sole source of carbon in the syn-

[1] a.d. $\sim PO_3^{=}$ refers to a high energy phosphate group in adenosine triphosphate or adenosine diphosphate. ad.H refers to adenylic acid. $H_2 \| O$ represents the photolytic splitting of water into hydrogen and oxygen.

thesis of complex organic structures, it has also been shown that heterotrophic bacteria can likewise assimilate CO_2.

1. *Formation of organic compounds by reduction of CO_2:* One of the simplest reactions of CO_2 in bacterial metabolism is its reduction to formic acid by H_2 in the presence of the enzyme hydrogenlyase (155).

In more complex systems, CO_2 is reduced to a one-carbon organic compound with the simultaneous oxidation of other metabolites (1). Thus, it has been shown by isotopic studies that *Methano-bacterium omelianski* is capable of oxidizing methanol or ethanol to their corresponding fatty acids with the simultaneous reduction of CO_2 to methane. In a similar manner, *Methanosarcina methanica* can oxidize fatty acids, derived from methanol or other alcohols, to CO_2 with the concurrent reduction of CO_2 to methane.

2. *Carbon to carbon addition by C_3-C_1 union (Wood-Werkman reaction):* The first indication that heterotrophic organisms can utilize CO_2 with the formation of carbon-to-carbon bonds came from the classical work of Wood and Werkman (148) who, using chemical methods only, were able to demonstrate an overall uptake of CO_2 by propionic acid bacteria during the fermentation of glycerol to succinic, propionic and acetic acids. By refining the methods of analysis, they were able to demonstrate that one molecule of CO_2 disappeared for every mole by succinic acid formed (149).

On the basis of these results, these investigators postulated that a one-carbon compound (CO_2) reacted with a three-carbon compound by C_3 to C_1 union to form the four-carbon dicarboxylic acid, succinic acid. Pyruvic acid was suggested as the intermediate with which CO_2 reacted.

a. *Studies of the Wood-Werkman reaction in vivo:* The introduction of the use of isotopes greatly facilitated the development of the work initiated by Wood and Werkman. Much of their earlier chemical work has been confirmed by the use of isotopes (30, 150, 151). Isotopic carbon was fixed in the carboxyl groups of succinic acid during its formation by the propionic acid bacteria (152). It was also unexpectedly discovered that during glycerol or glucose fermentation, the propionic acid formed likewise contained carboxyl isotopic carbon (151, 29, 28, 153). It has been suggested that the formation of succinate from pyruvate occurs through the following series of reactions:

$$\text{pyruvic acid} + CO_2 \leftrightharpoons \text{oxalacetic} \leftrightharpoons \text{malic} \leftrightharpoons \text{fumaric} \leftrightharpoons \text{succinic} \qquad (7)$$

Each of the reactions from oxalacetic to succinic acids has been demonstrated by Krebs and Eggleston (74) to occur reversibly in propionic acid bacteria.

Further evidence for the participation of malic and fumaric acids as intermediates in this reaction has been furnished by Nishina, Endo and Nakayama (100). They have succeeded in isolating crystalline derivatives of these acids from preparations of *Bact. coli communis* fermenting glycerol and glucose in the presence of radioactive CO_2. When ammonium chloride was present during the fermentation, radioactive aspartic acid could be isolated, thus indicating the probability of oxalacetic acid as an intermediate.

The steps concerned with the formation of propionic acid containing carboxyl

isotopic carbon derived from isotopic CO_2 are less certain, however. The suggestion was originally made by Carson et al. (30) and by Krebs and Eggleston (74) that isotopic propionic acid was formed by the reduction of isotopic pyruvic acid. This pyruvic acid was considered to be formed by a reversal of the reactions leading to the formation of the symmetrical isotopic succinic acid as shown in equation (7). Werkman and Wood (145) have objected to this scheme on the grounds that according to their data every molecule of propionic acid formed would have had to originate from pyruvic acid which had first undergone this series of side reactions from oxalacetic to succinic acids. They believe that it is improbable that this series of reactions takes place so much faster than the direct reduction of pyruvic acid to propionic acid. Werkman and Wood have suggested that, instead, propionic acid may arise from the splitting of succinic acid into propionic acid and CO_2. They are aware, however, that very little evidence is available for such an interpretation; for, as shown by Carson, Foster, Ruben and Barker (28), the addition of isotopic succinic acid to a preparation of propionic acid bacteria, fermenting glycerol in the absence of isotopic CO_2, results in the formation of very little isotopic propionic acid. Conversely, upon adding isotopic propionic acid to such a fermentation, only small amounts of isotopic succinic acid could be recovered.

The metabolism of CO_2 in several microorganisms other than the propionic acid bacteria has been studied (151, 120). Krebs (70) has made a tabulated summary of some of the microorganisms whose metabolism has been studied with the use of isotopic carbon. In general, organisms which during fermentation formed the dicarboxylic acid, succinic or the monocarboxylic acid, lactic or propionic, fixed isotopic carbon of CO_2. This isotopic carbon without fail was always located in the carboxyl carbon atom. At present, compounds of two of more carbon atoms synthesized by heterotrophic organisms and containing isotopic carbon originating from isotopic CO_2 are generally considered to arise by way of the Wood and Werkman reaction:

$$CH_3COCOOH + CO_2 \rightarrow HOOCCH_2COCOOH \qquad (8)$$

There are, however, notable exceptions to this generalization. The addition of a two-carbon compound to a one-carbon compound originating from CO_2, (equation (3)), has been established by Lipmann and Tuttle (83) and Utter, Lipmann and Werkman (135). The possible significance of this reaction in metabolism will be discussed in a later section.

The early finding that heavy carbon occasionally was found in the carboxyl position of acetic acid isolated from fermentations proceeding in the presence of heavy CO_2 indicated that CO_2 incorporation might occur by some mechanism hitherto undescribed (120). Recently, however, Werkman and co-workers (119, 64) have been able to account for the formation of isotopic acetic acid on the basis of the Wood and Werkman reaction. They considered the two possibilities that acetic acid may arise either by the oxidative decarboxylation of pyruvic acid or by the symmetrical reductive fission of succinic acid. In the former case, acetic acid should contain no isotope originating from CO_2 of the medium; in

the latter case, however, acetic acid should contain isotopic carbon in the carboxyl position since succinic acid is formed, in part at least, from pyruvic acid and isotopic CO_2. Their evidence indicates that both reactions occur.

It is also known that succinate may be formed by many organisms without involving the carbon dioxide fixation reaction (120). For example, the concentration of isotopic carbon in succinic acid isolated from the fermentations *Staphylococcus candidus*, *Streptococcus paracitrovorus* and *Proteus vulgaris* was below the theoretical value expected if one assumes that the entire amount of succinate is formed by the Wood and Werkman carboxylation reaction. These results coupled with other evidence give weight to Wood and Werkman's belief in the presence, in certain bacterial systems, of the reversible enzyme reaction:

$$\text{succinate} + 2\,H \rightleftharpoons 2\,\text{acetate} \qquad (9)$$

Recently, with the aid of two isotopic isomers of acetic acid and with isotopic succinate, Werkman and co-workers have been able to demonstrate the presence of this reaction in at least two species of bacteria. Slade and Werkman (119) have shown that suspensions of *Aerobacter indologenes* convert carboxyl labelled acetate into isotopic succinate and vice versa. In both experiments, the isotope was located in the carboxyl position of the organic products of reaction. When acetic acid was used in which the isotopic carbon was distributed equally in the molecule, the succinate obtained contained equal amounts of isotope in the methylene and carboxyl carbons. This experiment demonstrated that the isotope in the succinic acid originated directly from acetic acid rather than indirectly from CO_2 of the oxidized acetate since the carboxylation reaction did not yield products containing isotopic methylene carbon. Approximately fifty per cent of the acetate formed by *Aerobacter* fermenting glucose arose by the fission of succinic acid and thirteen per cent of the total succinate by condensation of acetic acid. Thus acetate condensation and succinate fission are reactions which possess quantitative significance.

Recently Kalnitsky and Werkman (61, 62) have obtained a cell-free extract of *Escherichia coli* which is capable of dissimilating pyruvate anaerobically with the formation of acetic, formic, lactic, and succinic acids. When fermentations were carried out in the presence of isotopic carbon dioxide, the last three acids contained isotopic carbon in their carboxyl groups thus demonstrating that the CO_2 assimilating enzymes were present in this cell-free extract. The concentration of isotope in the succinic acid was too low, however, to account for the synthesis of succinic acid solely by the Wood and Werkman reaction. By experiments with acetate containing isotopic carboxyl carbon, Kalnitsky, Wood and Werkman (64) were able to demonstrate that a significant part of the succinic acid arose by another path—i.e., the condensation of acetic acid.

The rôle of CO_2 in metabolism has been studied in many unicellular organisms other than bacteria. The molds, *Rhizopus nigricans* and *Aspergilus niger* ferment glucose and sucrose in the presence of $C^{11}O_2$ with the formation of isotopic fumaric and citric acids, respectively (53). In both cases, isotopic carbon was found only in the carboxyl groups. van Niel et al. (99) have shown that the

holotrichous ciliate, *Tetrahymena geleii*, fermenting sugar anaerobically in the presence of radioactive CO_2, fixes carbon in the carboxyl position of the succinic acid formed. This protozoan possesses one of the most active mechanisms for the assimilation of CO_2 of those systems studied.

The studies with isotopic CO_2 *in vivo* have demonstrated that the assimilation of CO_2 into organic molecules is a reaction of major importance in a variety of non-photosynthetic organisms. Much of the data obtained *in vivo* has been consistent with the hypothesis that CO_2 assimilation proceeds by way of the Wood-Werkman reaction. In order to learn more concerning the details of this reaction, experiments have been carried out with extracts of various tissues.

b. *Studies of the Wood-Werkman reaction with extracts:* The enzyme systems involved in the Wood and Werkman carboxylation reaction have been studied by Werkman and co-workers and by Evans and his group. Krampitz and Werkman (68) have isolated an enzyme from a preparation of *Micrococcus lysodeikticus* which decarboxylates oxalacetate to pyruvate in the presence of magnesium ions. Cocarboxylase was not needed for the reaction. Although the formation of oxalacetate from pyruvate could not be demonstrated by this bacterial preparation, there is evidence that it is involved in the pyruvate carboxylation reaction. When oxalacetate, the enzyme and magnesium ions were allowed to react in a medium containing $C^{13}O_2$ until approximately fifty per cent of the oxalacetate disappeared, C^{13} was found in the remaining oxalacetate in the carboxyl carbon α to the methylene group (145, 69).

The over-all formation of oxalacetate from pyruvate and CO_2 by extracts of *E. coli* has been recently demonstrated by non-isotopic experiments of Kalnitsky and Werkman (63).

The rôle of phosphate in the Wood-Werkman reaction is at present unknown. Krampitz et al. (69) believe that oxalacetate is in equilibrium with some three-carbon acid which is more readily carboxylated than pyruvic acid. They believe that the intermediate may be a phosphorylated derivative of pyruvic acid, not necessarily phosphopyruvic acid. "Oxalacetate" obtained during fumarate oxidation by an enzyme preparation of *Micrococcus lysodeikticus* was considerably enriched with isotopic carbon. Wood and Werkman consider that the "oxalacetate" formed by fumarate oxidation is more active, physiologically, than synthetic oxalacetate and that it undergoes decarboxylation to form the active carbon dioxide assimilating intermediate with greater facility.

The crude enzyme preparation from *Micrococcus lysodeikticus* contains no enzyme which is capable of distributing the assimilated heavy carbon of oxalacetate equally on both carboxyl carbon atoms. This is in contrast to the enzyme system prepared from pigeon liver by Evans, Vennesland and Slotin (49, 50). These investigators have obtained an enzyme preparation which can fix radioactive CO_2 in the presence of pyruvic acid through a series of reactions which result in the formation of radioactive pyruvic acid. They postulate that carboxyl isotopic pyruvate is formed by a series of reactions involving carboxylation of pyruvic acid to form carboxyl isotopic oxalacetate, distribution of the isotope between both carboxyl groups and decarboxylation of this dicar-

boxylic acid to pyruvate. Practically all of the CO_2 assimilated was present in the pyruvic acid isolated. When fumarate or malate as well as pyruvate was added to the enzyme preparation in the presence of radioactive bicarbonate, a considerably greater amount of radioactivity was fixed than when pyruvate alone was the substrate. The enzyme preparation contained an active fumarase system. For every mole of fumarate or malate that disappeared, one mole of CO_2 was produced and one mole of lactic acid was formed. The reaction has been divided into the following steps:

$$\left.\begin{array}{l}\text{fumarate} \rightleftharpoons \text{malate} \\ \text{malate} + \text{pyruvate} \rightleftharpoons \text{lactate} + \text{oxalacetate} \\ \text{oxalacetate} \rightleftharpoons C^{11}O_2 + \text{pyruvate}\end{array}\right\} \quad (10)$$

Only 15 per cent of the total radioactivity assimilated was present in the pyruvate isolated. The remainder was found in the filtrate and, in view of the lack of other intermediates in solution, was considered to be present in the lactate, malate and fumarate molecules. There is evidence that the dialyzed enzyme requires manganese ions and perhaps cozymase for reactivation. In contrast to the oxalacetate decarboxylase of bacterial origin, magnesium ions did not reactivate the enzyme prepared from pigeon liver.

3. *Carbon to carbon addition by C_2-C_1 union:* At present, with certain exceptions, the Wood and Werkman reaction accounts for the incorporation of CO_2 into organic molecules by heterotrophic systems. There have been attempts, however, to demonstrate the assimilation of CO_2 through addition to acetaldehyde according to the reaction:

$$CO_2 + CH_3CHO \rightarrow CH_3COCOOH \quad (11)$$

but so far, no substantial evidence exists demonstrating that this reaction, an example of C_2-C_1 union, occurs to a significant extent. Evans (46) has found no exchange of the carboxyl carbon of pyruvate and CO_2 in the presence of carboxylase even when the pressure of carbon dioxide was as high as 300 atmospheres. Carson et al. (31), in a similar type of experiment, could likewise find no significant amounts of radioactive carbon in the pyruvate isolated. In order to test whether, *in vivo*, any labelled pyruvate formed could be transformed or removed from the enzyme surface by a rapid reaction competing with the decarboxylation, Carson et al. have used the bacteria, *Streptococcus lactis*. This species dissimilates pyruvic acid into lactic acid, acetic acid and CO_2. At least two separate reactions are involved:

$$CH_3COCOOH + H_2O \rightarrow CH_3COOH + CO_2 + 2H^+ + 2e^- \quad (12)$$

$$CH_3COCOOH + 2H^+ + 2e^- \leftrightharpoons CH_3CHOHCOOH \quad (13)$$

Any radioactive pyruvate formed by the reversal of reaction (12) should be trapped as lactic acid by reaction (13). However, during the course of reaction no radioactive lactic acid was formed.

In spite of the failure to demonstrate the reversibility of the pyruvic acid decarboxylase reaction, Utter, Lipmann and Werkman (135) have recently reported that pyruvic acid may be formed from acetyl phosphate and formic acid (C_2-C_1 union). After incubating isotopic formic acid and non-isotopic pyruvate with extracts of *Escherichia coli*, both reactants were isolated and found to contain the same C^{13} concentrations in their carboxyl positions showing that an equilibrium had been reached. Other experiments by Lipmann and Tuttle (83) have demonstrated that acetyl phosphate is a component of this equilibrium. Since, at the end of incubation, the CO_2 had a much lower concentration of isotope C^{13} than did the formic acid and since the pyruvate had a high concentration of isotope, it may be concluded that the pyruvate had not been formed from a reaction of CO_2 and some other pyruvate precursor.

Although H_2 and CO_2 were not in equilibrium with formic acid in the bacterial extracts, such an equilibrium can apparently exist in intact cells. This is indicated by the formation of pyruvic acid from acetyl phosphate and CO_2 *in vivo*.

There has been an attempt to assess the relative importance of the C_3-C_1 and C_2-C_1 reactions in accounting for the assimilation of CO_2 into organic molecules during certain bacterial reactions. Slade et al. (120) have reported the formation of lactic acid with fixed carbon in the carboxyl group by *Clostridium welchii* and by *Clostridium acetobutylicum*. In spite of the formation of large amounts of lactic acid, the dicarboxylic acids were not synthesized to an appreciable extent. The lack of formation of dicarboxylic acids has raised doubt about their importance as intermediates of CO_2 fixation in lactic acid (via reaction (7)). In a further study of this problem, Brown, Wood and Werkman (22) have measured the rate of metabolism of succinic, fumaric, malic and oxalacetic acids by preparations of *Clostridium butylicum*. They have found that these acids are not metabolized at a sufficient rate to warrant their inclusion as intermediates of CO_2 fixation in lactic acid. Instead, they believe that C_2-C_1 fixation may account for their results.

The reversibility of the "phosphoroclastic" reaction is interesting because it demonstrates a second mechanism in addition to the Wood-Werkman reaction by which carbon dioxide may be incorporated into organic compounds. Although the C_2-C_1 reaction undoubtedly has significance in bacterial metabolism, there is as yet no information concerning the importance of a similar reaction in mammalian metabolism. Considerable evidence has accumulated pointing to the importance of the Wood-Werkman reaction in the metabolism of carbon compounds in mammalian tissues, however. This will be discussed in the succeeding section.

METABOLISM IN HIGHER FORMS: The introduction into biological work of isotopic carbon (C^{11} and C^{13}) has made possible new studies of mammalian intermediary metabolism both *in vitro* and *in vivo*. Although deuterium is invaluable in the study of fat and protein metabolism, its use is limited when applied to the study of carbohydrate reactions since deuterium is usually lost from carbohydrates and carbohydrate precursors by oxidation or by non-specific exchange reactions. During the study of carbohydrate metabolism *in vivo*, it became

evident that isotopic carbohydrate precursors are diluted by tissue constituents in much the same manner as are dietary proteins and fats. Use has been made of isotopic carbon to study the interaction of these carbohydrate precursors with materials of the diluting metabolic pool.

During the last eighty years, several theories (141, 102) have been proposed to explain the relationship between the foodstuffs of the diet and the protoplasmic and structural materials of the body. One of the most important of these theories was developed by Folin (51). Because of the constancy of excretion of certain waste products of metabolism—e.g., creatinine—and the variability of others—e.g., urea—Folin believed that there were two distinct kinds of protein metabolism occurring simultaneously which produced two different sets of waste products. One type of metabolism, resulting from "wear and tear" of the biological "machinery," was characterized by the constant rate at which its products were formed and was referred to as tissue, or endogenous metabolism; the other, resulting from the introduction of foodstuffs into the organism, was characterized by the variability of the excretion of its products and was called intermediate, or exogenous metabolism. According to Folin, there was little interaction between the components of these two biological systems. This theory of protein metabolism was widely accepted until recent years. The work of Whipple and Madden (89) and of Borsook and Keighley (16) raised questions regarding its correctness and their work has been supplemented and greatly extended by the experiments of Schoenheimer and his colleagues. Using the stable isotopes of nitrogen and hydrogen, they were able to show that special proteins (116, 117, 57) and creatinine (13, 15, 14), supposedly components of an isolated endogenous system, were derived to a considerable extent from dietary sources. From these and other experiments demonstrating the vast range of the interaction between foodstuffs and tissue constituents, Schoenheimer developed a new theory of metabolism. The essence of Schoenheimer's theory of the dynamic state of the body constituents has been succinctly stated by Clarke (32) as follows: "As a result of Schoenheimer's investigations, there has emerged a concept of metabolic 'regeneration,' wherein the central idea is the continual release and uptake of chemical substances by tissues to and from a circulating metabolic 'pool.' Coincident with these cyclic processes, there occur among the components of the pool multitudinous chemical reactions of which relatively few are concerned with elimination of waste products." Metabolic materials and reactions are, therefore, no longer to be thought of as segregated into exogenous and endogeous parts, but rather as a system whose components, whether in tissues or out, are in a state of dynamic equilibrium.

It is in the light of this theory of dynamic equilibrium that the subsequent work on metabolism of organic compounds will be presented.

A. *Relation of Metabolic Waste Products and Essential Foodstuffs to the Metabolic Pool:* Through the isotopic study of the nitrogen waste products of metabolism, the theory of a separate exogenous and endogenous metabolism was undermined. Folin considered that urea resulted solely from the breakdown of exogenous metabolites, whereas creatinine among other substances was thought to originate

at a constant rate solely from tissue constituents. In spite of the chemical relationship between creatine and creatinine, Folin (52) considered that the latter was not derived from the former, since he could not demonstrate an increase of urinary creatinine upon administering creatine. Although a relationship between the content of body creatine and urinary creatinine has been subsequently established by a number of investigators (95, 108, 2), it was not until the recent work of Bloch and Schoenheimer (13, 15, 14) that the metabolism of creatine was clarified. Creatinine is now known to originate solely from creatine. In the rat, approximately two per cent of the body content of creatine is converted into creatinine daily. The rate of creatine formation from dietary and tissue sources has likewise been found to occur at the same rate as its conversion to creatinine. The experiments of Bloch and Schoenheimer have demonstrated that as much as 25 to 30 per cent of the nitrogen in the newly formed creatine is derived from the isotopic components added to the stock diet containing casein. An additional part of the creatine formed is derived from the non-isotopic amino acid precursors in the diet as well as in the tissues of the animals. Creatine and indirectly creatinine are thus constantly formed from food as well as from tissue material.

Urea likewise is derived from both tissue and dietary constituents. Ammonia, as well as the amino groups of amino acids are ready sources of the nitrogen of urea (115). CO_2 has been shown by the isotopic method to play a rôle in urea formation. Liver slices in the presence of isotopic CO_2 produced urea containing high concentrations of carbon isotope (48, 107).

Whatever the biological factors may be that regulate the excretion of creatinine and permit the variability of urea excretion, they are not adequately explained by the existence of a separate endogenous and exogenous metabolism. Instead, the nitrogenous end-products of metabolism are formed from a common metabolic pool, the constituents of which have been derived from both the tissues and the diet.

Creatinine and urea are alike in that, once formed, they do not re-enter the metabolic reactions of the body (115). The administration of isotopic creatinine and urea results in the excretion of the compounds in the urine with unchanged isotope concentrations. Because creatinine and urea once formed do not enter into metabolic reactions, one may consider that they lie outside that body of metabolites which are known collectively as the "metabolic pool."

There is still another group of substances which may be placed in this same category. These are the essential foodstuffs such as the essential amino acids, essential unsaturated fatty acids, and probably many of the vitamins (4). Although during their metabolism these essential substances are undoubtedly catabolized into products which mix intimately with the large metabolic pool, they are apparently not regenerated from it. The two requirements of a metabolite before it can be considered as a part of the "metabolic pool" are that it be capable of formation from and contribute to the synthesis of other components of the pool. In other words, there exists among the components of the metabolic pool a state of dynamic equilibrium.

B. *Relation of Structural and Storage Materials to the Metabolic Pool:* The extent to which the structural and storage elements of the organism react with metabolites of the metabolic pool is a function of the rate at which these materials are synthesized and degraded in the animal. The structural or storage materials of the body are usually of high molecular weight and are often relatively insoluble in water. Consequently, one would not expect these substances to come into complete equilibrium with the components of the metabolic pool at the same rate as do the highly reactive, water-soluble, small molecular weight compounds. They will in time, however, reflect the composition of the metabolic pool from which they originate. The length of time required for equilibrium to occur depends upon the individual structural or storage material. Thus, mice forming fat from carbohydrate will resynthesize one-half of their total fat depots during a period of seven days (5). Consequently, if the animal maintains a static weight for seven days, one-half of the original amount of fat in the depot will have been degraded and contributed to the metabolic pool in the form of oxidizable intermediates. Liver fat is similarly synthesized and degraded seven times faster than is the depot fat (5); cholesterol turnover, however, is about one-third as fast (106).

It may be stated at once that, after glycogen, fat, and protein have been degraded into their simpler elements—i.e., pyruvic acid, fatty acids, and amino acids—the reactions involved in the further degradation of these substances into CO_2 and water have much in common (fig. 1).

The mechanism of oxidation of carbohydrates was the first to be investigated. Szent-Györgyi (132) and Stare and Baumann (125) found that the addition of catalytic amounts of certain dicarboxylic acids stimulated considerably the oxidative processes of isolated tissues. As a result, they assigned to these acids a rôle in the hydrogen transport mechanism of tissue oxidations. Later Krebs and Johnson (76), with the knowledge that citric acid as well as α-ketoglutaric acid acted as catalysts in oxidative reactions, postulated the formation of a six-carbon, tricarboxylic acid by the condensation of oxalacetic acid with a three-carbon carbohydrate intermediate accompanied by decarboxylation. The successive oxidation of citric acid, the postulated intermediate, to oxalacetic acid regenerated one of the reactants, thus explaining the catalytic effect of small amounts of dicarboxylic acids on carbohydrate oxidation. The net result was the oxidation of carbohydrate to carbon dioxide and water. This cycle of reactions was first called the "citric acid cycle." It is now more commonly referred to as the "tricarboxylic acid cycle."

Certain amino acids, alanine, glutamic and aspartic acids, are closely related to components of the tricarboxylic acid cycle. By transamination reactions first shown by Braunstein and Kritzmann (19), these three acids may be transformed into their respective keto acids: pyruvic, α-ketoglutaric and oxalacetic. For example:

$$CH_3\cdot CO\cdot COOH + HOOC\cdot CH_2\cdot CH_2\cdot CHNH_2\cdot COOH \rightleftharpoons CH_3\cdot CHNH_2\cdot COOH + HOOC\cdot CH_2\cdot CH_2\cdot CO\cdot COOH \quad (14)$$

Since these reactions are reversible, acids of the tricarboxylic acid cycle can either be formed from or contribute to the synthesis of certain amino acids. In addition to enzymes of the transamination system, there is an l-amino acid oxidase which reversibly converts glutamic acid into α-ketoglutaric acid and ammonia (44).

Recently it has been demonstrated that the organic acids of the tricarboxylic acid cycle are intermediates in the oxidation of fatty acids and ketone bodies (20, 21, 146, 26, 105).

The organic acids of the tricarboxylic acid cycle are consequently the meeting point of protein, fat and carbohydrate metabolism. Since the conversion of one foodstuff into another may involve reactions of the tricarboxylic acid cycle, a substance may be diluted by the components of the cycle during the course of the reaction. Even though the tricarboxylic acid cycle may not be directly concerned in a given reaction of the organism, the reacting substance may come into equilibrium with the components of the cycle before undergoing its primary transformation. Thus, the composition of the primary product of reaction will reflect the secondary circuitous equilibrium reactions involving the reactant.

1. *The tricarboxylic acid cycle and the rôle of CO_2:* The mechanism of the oxidative dissimilation of carbohydrates has been studied by Krebs and Eggleston (72) using pigeon breast muscle. As mentioned previously, they found that the presence of a small amount of dicarboxylic acid was an essential component of the system which oxidizes pyruvic acid. Since only small amounts of oxalacetic acid are necessary to effect the oxidation of large amounts of pyruvic acid, oxalacetate must be continually regenerated by the cyclic reactions by which pyruvic acid is oxidized. The addition of malonic acid, a specific inhibitor of the succinoxidase system, prevents the regeneration of oxalacetic acid and causes the accumulation of the products of the condensation of pyruvic and oxalacetic acids—i.e., citric, α-ketoglutaric and succinic acids. When a tissue, poisoned with malonate, uses up its endogenous sources of oxalacetate, the oxidation of pyruvate stops and can only be started again by the addition of oxalacetate or its equivalent. The amount of pyruvate disappearing under these circumstances is equivalent to the amount of oxalacetate added to the tissue.

In contrast to pigeon muscle mince, Evans (45) has found that pigeon liver mince poisoned with malonate does not require the addition of dicarboxylic acids to oxidize pyruvic acid. Apparently, pigeon liver mince is capable of synthesizing its own dicarboxylic acids from pyruvate. Krebs and Eggleston (73), from indirect evidence, suggested that the dicarboxylic acids arose from pyruvate by means of the carboxylation reaction of Wood and Werkman (equation (8)). In their experiments with normal pigeon liver mince, a decrease of utilization of pyruvate and consequently of dicarboxylic acid production was observed when bicarbonate was excluded from the incubating medium. Thiamine was also demonstrated to play a rôle in pyruvate dissimilation. The rate of pyruvate dissimilation was decreased in liver suspensions of thiamine-deficient pigeons containing an adequate amount of bicarbonate. The addition of thiamine to these suspensions increased the rate of production of dicarboxylic acids several fold.

Direct evidence for the utilization of carbon dioxide by rat liver was obtained by Ruben and Kamen (111) using radioactive carbon. Although assimilation of carbon dioxide was demonstrated in their experiments, no specific reaction products containing radioactive carbon were isolated.

However, Evans and Slotin (47), extending their earlier experiments, isolated radioactive α-ketoglutaric acid formed by pigeon liver mince from pyruvate in

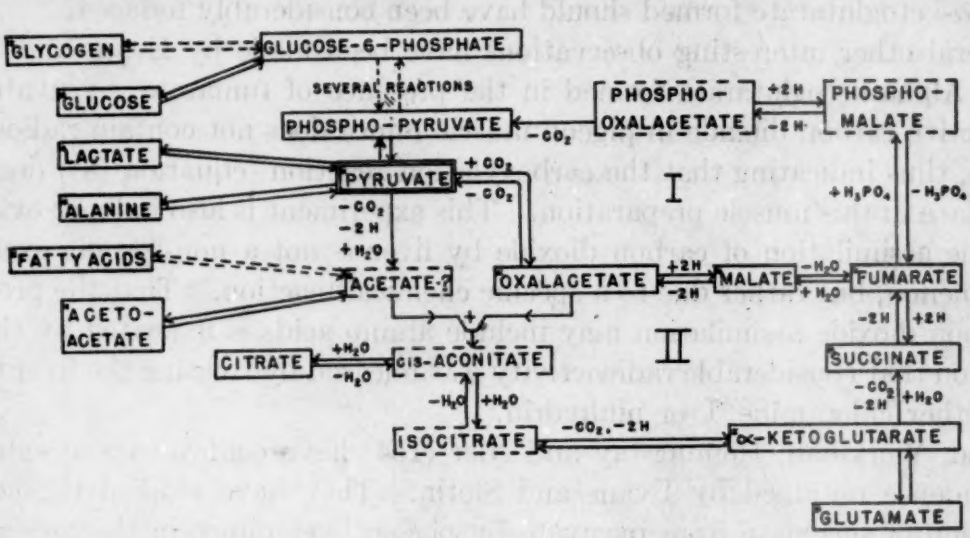

Fig. 1. Diagram illustrating reactions involving examples from carbohydrate, protein and fat metabolic pools.

Cycle I: Gives hypothetical reactions of pyruvate and CO_2 in formation of phosphopyruvate.

Cycle II: Gives reactions of tricarboxylic acid cycle in oxidation of pyruvate to $3CO_2 + 2H_2O$.

Compounds marked * have been studied by isotopic analysis.

Broken boxes designate substances of hypothetical significance.

Broken lines designate reactions omitted.

Formulas of acids of cycles I and II are to be found in figure 3 except:

Succinic: $HOOC \cdot CH_2 \cdot CH_2 \cdot COOH$

Cis-aconitic: $HOOC \cdot CH{:}C \cdot CH_2 \cdot COOH$
$\quad\quad\quad

contain radioactive carbon in either of its carboxyl groups. However, it did not. The observations of Evans et al. indicate, therefore, that citric acid cannot be a direct intermediate in the formation of α-ketoglutaric acid from pyruvic acid by liver mince. Furthermore, the specific radioactivity of the α-ketoglutarate formed was not altered by the addition of citrate to the medium. If citrate were an intermediate in the dissimilation of pyruvate, the specific radioactivity of the α-ketoglutarate formed should have been considerably reduced.

Several other interesting observations have been made by Evans and Slotin (47). Alpha-ketoglutarate formed in the presence of fumarate, pyruvate and radioactive carbon dioxide in pigeon muscle mince does not contain radioactive carbon, thus indicating that the carboxylation reaction (equation (8)) does not take place in this muscle preparation. This experiment is also indirect evidence that the assimilation of carbon dioxide by liver is not a non-specific exchange phenomenon, but rather due to a specific chemical reaction. That the products of carbon dioxide assimilation may include amino acids is indicated by the observation that considerable radioactivity was released by treating the liver mince with either chloramine T or ninhydrin.

Wood, Werkman, Hemingway and Nier (154) have confirmed and extended the evidence obtained by Evans and Slotin. They have studied the aerobic formation of succinate from pyruvate by pigeon liver mince in the presence of malonate and its synthesis anaerobically in the absence of inhibitor. In the former case, the succinate isolated did not contain isotopic carbon derived from the isotopic CO_2 present in the medium. As in the experiments of Evans and Slotin, the α-ketoglutarate contained excess isotopic carbon in the carboxyl group α to the keto group. Fumarate and malate both contained isotopic carbon in the carboxyl position. Since succinate contained no isotopic carbon, it was considered to arise solely from the oxidative decarboxylation of α-ketoglutarate. This experiment demonstrated conclusively that malonate in the concentrations used does completely inhibit succinic dehydrogenase and hence prevents the formation of succinic acid by the reduction of fumaric acid. As expected, in anaerobic experiments in the absence of malonate, the succinate as well as the fumarate and malate contained excess isotopic carbon in the carboxyl position.

These experiments have demonstrated that there are at least two pathways for the formation of succinate from pyruvate by pigeon liver mince (fig. 1). By one path, succinate is formed by reductive steps from oxalacetate, the latter being synthesized from pyruvate and carbon dioxide according to the Wood and Werkman reaction. The last step in this series—i.e., fumarate $\leftrightharpoons$ succinate—is completely inhibited by malonate. The other path for succinate formation is aerobic. There are no enzymes in this series of oxidative reactions which are specifically inhibited by malonate. According to Wood at al., the first step of this series is an oxidative condensation of oxalacetate and pyruvate to form cisaconitate. This latter compound is hydrated to isocitrate which is oxidized and decarboxylated to α-ketoglutarate. The place of isocitrate and cisaconitate in this scheme is at present tentative. Evans and Slotin believe that cisaconitate cannot be the intermediate in this series of reactions since citrate, isocitrate and

cisaconitate are in equilibrium with each other. Krebs (71), on the other hand, postulates that the rate of conversion of cisaconitate to α-ketoglutarate is so much faster than its conversion to citrate that equilibrium between cisaconitate and citrate never occurs, at least, in pigeon liver.

Ochoa and Weisz-Tabori have recently demonstrated the presence of oxalosuccinic carboxylase in pig heart preparations (100a). This observation has led Ochoa to the conclusion that the enzyme system composed of aconitase, isocitric dehydrogenase and oxalosuccinic carboxylase may provide another important manner in which CO_2 is utilized (100b).

Thus, it may be stated that although the tricarboxylic acid cycle provides a reasonable and logical description of the chemical reactions concerned with the conversion of pyruvate to CO_2 and water, and although there exists considerable evidence of an indirect nature pointing to its significance *in vitro*, direct proof for its correctness in detail and the magnitude of the rôle it plays *in vivo* await further data.

2. *Glycogen synthesis and glucose metabolism.* There are two fortunate factors which contribute greatly to the isotopic study of the synthesis of liver glycogen from its precursors. One of these is the rapidity with which glycogen formation takes place in the liver after the administration of a variety of glycogenic substances. The other is the fact that the preformed stores of liver glycogen can be reduced to a very low amount by fasting, so that the composition of the glycogen isolated after the administration of an isotopic glycogenic substance directly reflects the composition of the compounds from which it originates.

a. *Reactions of isotopic CO_2 and lactate in liver glycogen synthesis:* During the course of experiments, investigating the sources of glycogen carbon, Solomon, Vennesland, Klemperer, Buchanan and Hastings (121) observed that starved rats fed non-isotopic lactic acid and injected intraperitoneally with isotopic sodium bicarbonate, ($NaHC^{11}O_3$), had an appreciable quantity of radioactive carbon in the glycogen which was formed in the liver.[2] From the specific radioactivity of the expired CO_2, and the specific radioactivity of the newly formed glycogen carbon, the proportion of this glycogen carbon which had been CO_2 in the body was estimated and found to average 11 per cent. In three out of the seven experiments reported, the proportion was 13.9, 14.2, and 15.6 per cent. These results were interpreted as being consistent with the reactions suggested

[2] It should be recorded that, at a conference of the Harvard investigators early in the fall of 1939, our former colleague, Dr. Birgit Vennesland, had, with infallible feminine intuition, suggested the possibility of CO_2 incorporation in glycogen, and proposed an experiment with radioactive CO_2 and non-radioactive lactate. This suggestion seemed fantastic to the undersigned and it was not carried out until the spring of 1940. And even then, it was performed as a control experiment to insure the adequacy of the experimental procedures employed when radioactive lactic acid was fed.

Finding radioactivity in the liver glycogen whose source was CO_2 continued to be viewed with skepticism until Doctor Vennesland succeeded, after several disappointments, in winning the race against the five hour deadline set by the rapidly decaying C^{11}, and demonstrated the presence of the radioactivity in the osazone isolated from hydrolyzed glycogen (137). A.B.H.

by Lipmann (82) and Kalckar (60) for the formation of phosphopyruvate from dicarboxylic acids (fig. 2). According to this hypothesis, the reactions tracing the fate of isotopic carbon atoms starting with the Wood-Werkman reaction (pyruvate + CO_2) are given in figure 2. It is to be noted that these reactions

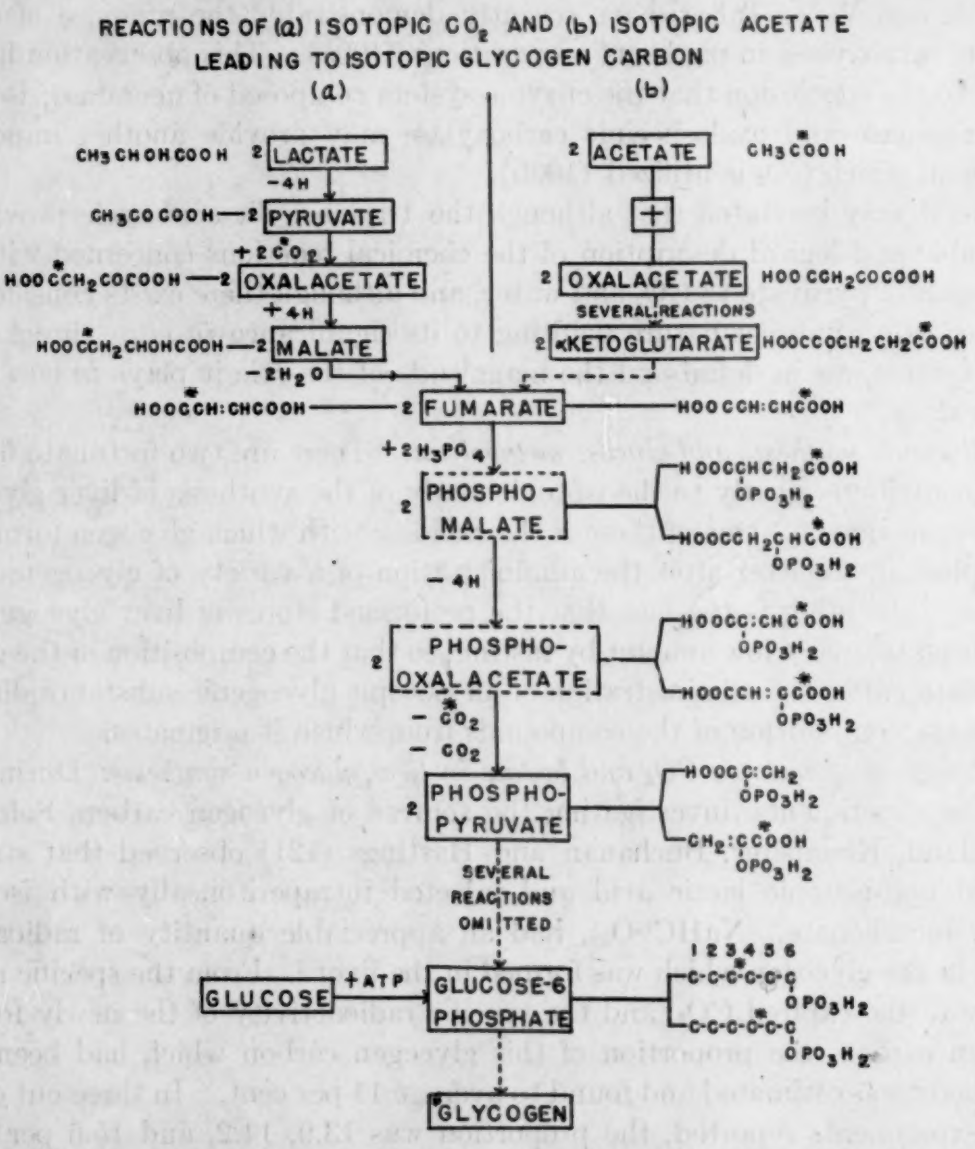

Fig. 2. The positions of isotopic carbon in glycogen intermediates are indicated by *.

Reactions proposed for incorporation of isotopic carbon in liver glycogen *in vivo* (a) after C^*O_2 and lactate administration (Ref. 47, 154, 121); after C^*O_2 and glucose administration (Ref. 137, 147); (b) after isotopic acetate, CH_3C^*OOH, and glucose (Ref. 81, 26). Reactions of C^*O_2 are shown at left; reactions of isotopic acetate at right. Isotopic carbons from both sources are indicated as having identical positions from fumarate to glycogen.

Broken boxes designate hypothetical compounds; broken lines designate omitted reactions. Only the carbons of glucose-6-phosphate have been indicated.

would predict that one-sixth of the carbon atoms of glycogen were derived from CO_2 carbon.

Recent experiments with non-isotopic glucose and with isotopic $C^{13}O_2$ by Wood, Lifson and Lorber (147) have proved that the CO_2 carbon occupies position 3 or 4 in the glucose derived from the newly formed glycogen.

From work recently reported by Lardy and Ziegler (77), it would appear that direct phosphorylation of pyruvate to form phosphopyruvate can occur in extracts of rat muscle in the presence of potassium ions. Previous attempts to demonstrate the direct phosphorylation of pyruvate had been unsuccessful (94) and had led to the hypothesis of Lipmann and of Kalckar referred to above. The observations of Lardy and Ziegler taken together with the experiments of Evans, Vennesland and Slotin (50) and of Wood, Vennesland and Evans (147a) present an alternate pathway whereby isotopic carbon from C^*O_2 could participate in glycogen formation. (See fig. 1 and the following reactions.)

$$\left.\begin{array}{l}\text{pyruvate} + C^*O_2 \rightleftharpoons \text{oxalacetate}^* \\ \text{oxalacetate}^* \rightleftharpoons \text{pyruvate}^* + CO_2 \\ \text{pyruvate}^* + \text{ATP} \rightleftharpoons \text{phosphopyruvate}^* + \text{ADP}\end{array}\right\} \quad (15)$$

Until further work on this question has been carried out, it is not possible to evaluate the relative biological importance of the two ways whereby phosphopyruvate may be formed *in vivo*.

b. *Reactions of isotopic lactate in liver glycogen synthesis:* Although the synthesis of liver glycogen is known to be increased by lactate administration (34), the extent to which the lactate molecules administered comprise the glucose units of the glycogen was unknown prior to the use of isotopic lactate. Experiments have now been reported using isotopic lactate with radioactive carboxyl carbon ($CH_3CHOHC^{11}OOH$) (37, 33), and with radioactive α or β carbons ($CH_3C^{11}HOHCOOH$ or $C^{11}H_3CHOHCOOH$), (136) which clarify this question.

It was found that in rats fed isotopic α-β carbon lactate an average of 16 per cent of the glycogen formed was derived from three-carbon lactate units after allowing for the radioactivity due to CO_2 incorporation. This may be interpreted as indicating a six-fold dilution of the lactate molecules in the effective carbohydrate metabolic pool. The reactions which would account for the presence and positions of the radioactive carbons in the glucose units of the glycogen are indicated in the left hand column of figure 3.

These reactions would predict the presence of isotopic carbon in positions 1–6, 1–5, 2–6, or 2–5 of the glucose unit. The correctness of this prediction of the positions of the isotopic carbon has not been proven.

In the experiments with isotopic carboxyl lactate, the specific radioactivity of the glycogen formed was considerably less than in the experiments with isotopic α-β carbon lactate. It would appear from the data presented that an average of only 6 per cent of the glycogen was derived from three-carbon lactate units after allowing for the radioactivity due to CO_2 incorporation. However, when the reactions concerned in the conversion of isotopic carboxyl lactate are reviewed (fig. 3, right-hand column), it is seen that only half as much isotope per glucose unit would be expected as in the experiments with α-β lactate. This is due to the loss of radioactive carbon as CO_2 from half the original three-carbon lactate units. Since the reactions listed in figure 3 predict that twice as much isotopic carbon is to be expected in glycogen derived from isotopic α-β carbon lactate as from isotopic carboxyl carbon lactate, and since 16 per cent and 6 per

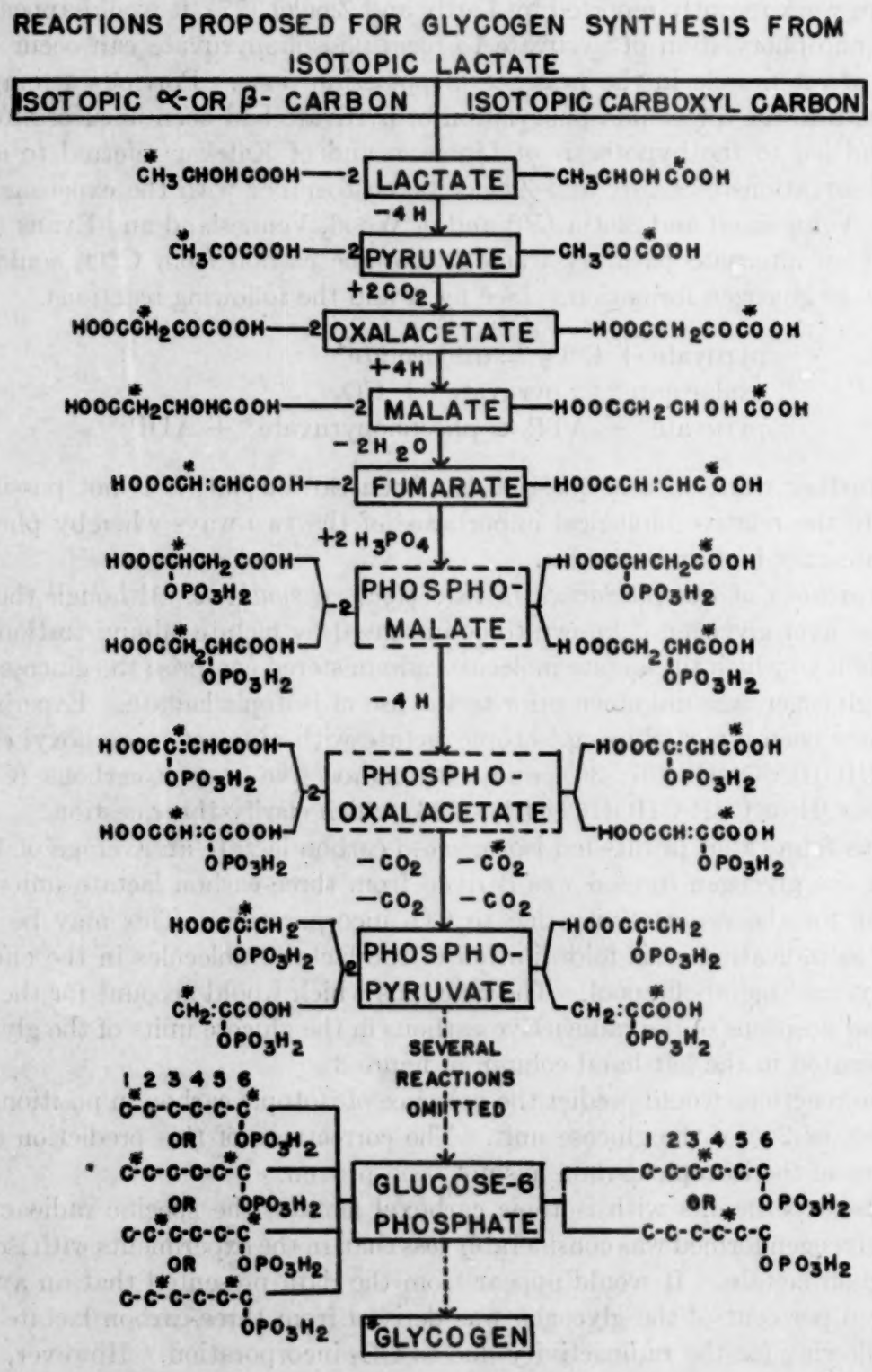

Fig. 3. The positions of isotopic carbon in glycogen intermediates are indicated by *. Reactions of isotopic β carbon lactate only shown at left. Reactions of isotopic α carbon lactate would give same end result and have been omitted (Ref. 136).

Reactions of isotopic carboxyl carbon shown at right (Ref. 33).

Broken boxes designate hypothetical compounds; broken lines designate omitted reactions. Only the carbons of glucose-6-phosphate molecules have been indicated.

cent, respectively, were found, support is given to the over-all theory of CO_2 assimilation during glycogen synthesis. The correctness of the details of reactions listed in figure 3 and the positions occupied by the isotopic carbon in the glucose units await further study.

c. *Reactions of isotopic CO_2 and glucose in liver glycogen synthesis:* Experiments have also been carried out in which glucose (137) was fed to a fasted animal and radioactive bicarbonate injected. Approximately 13 per cent of the carbon of the newly formed glycogen was found to have originated from carbon dioxide. This is in good agreement with the amount of carbon dioxide carbon entering glycogen during its formation from lactate.

During the synthesis of glycogen from glucose *in vitro* by rabbit liver slices, however, only small amounts of radioactive carbon dioxide were incorporated into the glycogen formed (137). This is in contrast to the results obtained *in vitro* with radioactive $NaHC^{11}O_3$ and non-isotopic pyruvate as substrate instead of glucose. Approximately 12 per cent of the glycogen formed *in vitro* from pyruvate by rabbit liver slices originated from carbon dioxide (24). In experiments, as yet unpublished (23) in which non-isotopic pyruvate was fed to rats and isotopic $NaHC^{11}O_3$ was injected, glycogen was formed, *in vivo*, which contained 14 per cent carbon originating from CO_2.

The agreement between the results obtained with pyruvate *in vivo* and *in vitro* and the disagreement obtained in the case of glucose is of interest in considering the metabolism of glucose in the body. According to the present theories of the formation of glycogen from glucose developed from the work by Cori and his co-workers using pure enzyme systems, no step would appear to require carbon dioxide incorporation (35).

$$\text{glucose} \rightarrow \text{glucose-6-phosphate} \rightarrow \text{glucose-1-phosphate} \rightarrow \text{glycogen} \qquad (16)$$

That this is probably the case during glycogenesis by a tissue system in which dilution and equilibrium phenomena play a minor rôle is indicated by the *in vitro* experiments with rabbit liver slices. In the intact animal, however, absorbed glucose is subject to many reactions which do not occur in the isolated system. The fact that the radioactive carbon of injected $NaHC^{11}O_3$ is incorporated into glycogen to the same extent after glucose feeding as after lactate and pyruvate feeding indicates that the glycogen deposited has originated in part, at least, from compounds which have undergone a carbon dioxide incorporation reaction. Since the glucose absorbed must have represented a significant portion of the final metabolic pool, it must be assumed that the ingested glucose comes into equilibrium very rapidly with the dicarboxylic acid pool before its conversion to glycogen. This may occur according to the reactions shown in figure 1. The reactions between glucose-6-phosphate and the dicarboxylic acids and pyruvate are assumed to occur much more rapidly than the conversion of glucose-6-phosphate to glycogen.

As stated in the preceding section, Wood, Lifson and Lorber (147), using C^{13}, have provided further impressive evidence for the validity of the above reactions for CO_2 assimilation into glycogen following the administration of glucose to

white rats. By chemical and bacteriological procedures, these investigators have found that C^{13} is located in the third or fourth carbon atoms of the glucose obtained from the liver glycogen of rats receiving glucose and isotopic sodium bicarbonate. These results would have been expected if glycogen is synthesized from three-carbon units by way of the dicarboxylic acids and the classical glycolytic scheme (fig. 2).

Although glucose plays an active rôle in the equilibrium reactions of the metabolic pool, glycogen is relatively inert in this respect. When radioactive bicarbonate was injected into a well-fed rat with large stores of liver glycogen, no radioactivity was found in the liver glycogen isolated after two and one-half hours. Thus, the incorporation of C_{11} in liver glycogen occurs to an appreciable extent only when the animal is actually making and depositing glycogen (121).

d. *Glycogen synthesis in muscle:* A study has also been made of glycogen synthesis in muscle tissue after feeding lactate, pyruvate or glucose and injecting $NaHC^{11}O_3$. After the administration of glucose and $NaHC^{11}O_3$, the muscle glycogen contained significant amounts of C^{11} (137). Such was not the case after feeding lactate or pyruvate (121, 23). Whether or not new muscle glycogen was formed in these experiments is not known.

On the other hand, Boxer and Stetten (18), using the presence in glycogen of deuterium, after the administration of D_2O, as a measure of glycogen synthesis, found an appreciable formation of muscle glycogen 4 hours after the administration of lactate to a fasted rat. The failure of C^{11} to appear in muscle glycogen after the administration of lactate may have been due to lack of glycogen synthesis in this tissue during the experiments cited above or to a difference in the mechanism of glycogen synthesis in muscle and liver. It is known, however, that CO_2 may be incorporated into glycogen during its synthesis by muscle tissue. This has been shown by the experiments of Lorber, Hemingway and Nier (84), who have studied the formation of muscle glycogen in an isolated cat heart perfused with a solution containing glucose or lactic acid and isotopic carbon dioxide. Regardless of the nature of the substrate, the glycogen isolated from the heart muscle contained an excess of heavy carbon. By assuming that each isotopic hexose unit of the glycogen contained only one isotopic carbon atom, Lorber et al. have calculated that, on an average, one-sixth of the total glycogen isolated after an experimental period of 1.5 hours was formed by reactions involving the incorporation of carbon dioxide.

It is entirely possible that two paths of glycogen synthesis from lactate occur, the one supplementary to the other. In certain tissues such as muscle, where glycogen synthesis from lactate does not take place at a rapid rate and where CO_2 assimilation reactions are not particularly marked, the predominant pathway of glycogen formation may be by the direct conversion of pyruvate to phosphopyruvate. In contrast, in liver tissue, where glycogen synthesis from lactate does take place at a rapid rate, glycogen formation involving the synthesis of dicarboxylic acids and the incorporation of CO_2 may be the primary pathway.

e. *Nutritional factors and glucose metabolism:* Ussing (134) demonstrated that stably-bound deuterium is incorporated into liver glycogen during its synthesis

from glucose by rats injected with deuterio-water. Since the direct conversion of glucose to glycogen according to equation (16) should not involve the incorporation of stably-bound deuterium into the glycogen molecule, Ussing's results have been interpreted as meaning that part of the glycogen was synthesized from carbon units smaller than glucose. From the reactions of figure 2, it would be theoretically possible for all of the hydrogen atoms of a three-carbon glycogen precursor such as lactic acid to be removed or replaced by deuterium during its conversion to glycogen. By feeding lactate to fasted rats injected with deuterio-water, Boxer and Stetten (18) have demonstrated that deuterium of deuterio-water is incorporated into practically all the positions of the glycogen molecule theoretically possible during its synthesis from this three-carbon precursor. In subsequent experiments in which glucose and deuterio-water were fed to animals and glycogen isolated under a variety of experimental conditions, these investigators have assumed that glycogen containing deuterium is formed by reactions involving small carbon chains, whereas glycogen containing no excess of deuterium results from the direct conversion of dietary glucose to glycogen by way of reactions of equation (16). By comparing the deuterium concentration of newly formed glycogen with the deuterium concentration of the body water in which it was formed, Boxer and Stetten have estimated in their experiments the proportion of glycogen synthesized by each of these two pathways.

When deuterio-water was administered to well-nourished animals on a high carbohydrate diet, the deuterium concentration of the liver glycogen gradually increased until it reached a constant level from 8 to 16 days after the beginning of the experiments (127). At this "saturation level" it was estimated that 43 per cent of the newly formed liver glycogen was derived by reactions which incorporated deuterium—i.e., by synthesis from precursors with carbon chains smaller than glucose. Only 3 per cent of the total glucose metabolized was converted to glycogen, demonstrating that the conversion of glucose to glycogen is not a major pathway of glucose metabolism. Under certain conditions, the metabolism of the animal on a high carbohydrate diet could be altered so that a greater percentage of newly formed glycogen originated from carbon chains smaller than glucose. These conditions were (a) the administration of adrenalin (129); (b) the administration of alloxan (128); (c) thiamine deficiency (17), or (d) fasting of the animal 24 hours prior to the feeding of glucose (18). In each of these experiments, the isolated glycogen contained a greater deuterium concentration than a sample isolated from a normal, well-nourished control. In contrast, the administration of glucose and insulin to a fasted animal resulted in the formation of glycogen predominantly by the union of hexose units. The results with adrenalin confirm earlier experiments of Cori and Cori (36) who have stated that glycogen deposited in the liver of the previously fasted rat in response to adrenalin was derived from blood lactate.

Deuterium of deuterio-water is also incorporated into fatty acids in constant amounts during their synthesis from glucose (115). Using this as a measure of fat synthesis from glucose, Stetten and Boxer have studied the rate of fat synthesis after the feeding of glucose to animals under some of the various experi-

mental conditions listed above. Thirty per cent of the daily glucose supplements of the diet of a well-nourished animal is converted into fatty acids (127). When these animals were poisoned with alloxan, thus producing an insulin deficiency, the rate of fatty acid synthesis was reduced to 5 per cent of the normal value (128). This blocking of fat synthesis in hypoinsulinism thus represents the cessation of a major pathway of carbohydrate utilization. No such depression of fatty acid synthesis was observed in the case of the phlorhizinized animals where the islet tissue is intact (129).

The rate of fatty acid synthesis from glucose was also greatly diminished in the thiamine-deficient animal (17). In another group of normal animals whose caloric intake was lowered to that of the thiamine-deficient group, the synthesis of fatty acids was likewise hindered. Boxer and Stetten, consequently, consider that the lack of synthesis of fatty acids in the thiamine-deficient animal resulted from inanition rather than from any specific effect of the thiamine deficiency on carbohydrate metabolism.

3. *Metabolism of fatty acids:* Although much has been learned concerning the metabolism of carbohydrates, there has been until recently but little conclusive experimental evidence available on the intermediates of fatty acid and ketone body oxidation. In the present section, this new evidence gained from isotopic studies will be presented and its interpretation in terms of current theories of fat metabolism will be discussed.

a. *Formation of ketone bodies:* The classical experiments of Knoop (66) and Dakin (38) were largely responsible for establishing the theory of successive β-oxidation of fatty acids. According to this theory, two-carbon compounds are successively split off from fatty acids and oxidized. Hurtley (58) in 1916 was the first to raise objection to this scheme of oxidation of fatty acids. He proposed the theory of multiple alternate oxidation to account for the absence of short chained fatty acids in the urine of diabetics excreting excessive amounts of ketone bodies. According to this hypothesis, approximately four ketones are produced by the oxidation of one sixteen-carbon fatty acid, instead of one which would be expected from the theory of successive β-oxidation. This point of view has been extended by work of Deuel (27), Stadie (122), Blixenkrone-Møller (9), Quastel (59) and their associates. Deuel and co-workers (27) have found that the administration of fatty acids with chains of four to ten carbon atoms resulted in the excretion of ketone bodies equal to or considerably more than the amount excreted in the urine after feeding equimolecular amounts of acetoacetic acid. Blixenkrone-Møller, measuring the ketone production and oxygen consumption of perfused diabetic livers, came to the conclusion that the low ratio of oxygen consumed to ketones produced could not be accounted for by the theory of β-oxidation. Stadie, Zapp, and Lukens (123) have extended Blixenkrone-Møller's experiments. By measurement of oxygen consumption, carbon dioxide and urea production of liver slices, they have been able to estimate the amount of oxygen that the tissue used in the production of the ketone bodies found. About 1.11 moles of ketone bodies were produced for every mole of oxygen disappearing. The theory of multiple alternate oxidation requires

a ratio of oxygen consumed to ketones produced of 1.25; the theory of β-oxidation, 6.5.

In spite of these excellent quantitative experiments, it has been generally recognized that β-oxidation does take place in the organism. Schoenheimer and Rittenberg (118) found that deuterio palmitic acid was deposited in the depots of animals fed deuterio stearic acid. Likewise, animals receiving deuterio palmitic acid were capable of synthesizing deuterio stearic acid (130). Such experiments provided substantial evidence in support of the theory of successive β-oxidation of fatty acids.

From these diverse and conflicting data has evolved a theory of fat oxidation which represents a compromise between the two theories referred to above. According to MacKay and co-workers (86), fatty acids undergo successive β-oxidation forming two-carbon compounds which immediately condense to produce a four-carbon ketone body. This hypothesis then embodies the mechanism of oxidation of fatty acids suggested by Knoop, but is more complete than the older theory since it accounts for the accumulation of ketone bodies during fat metabolism. Moreover, this theory explains certain findings which cannot be reconciled with the theory of multiple alternate oxidation—namely, the formation of more ketone bodies from hexanoic than from butyric acid (41) and the formation of ketone bodies from valeric acid, three carbon atoms of which are known to form glycogen (79, 87).

The recent study of the oxidation of carboxyl labelled fatty acids to acetoacetate by rat liver slices has rendered considerable support to MacKay's theory of β-oxidation-condensation. Weinhouse, Medes and Floyd (143) have found that acetoacetate formed from the oxidation of carboxyl labelled octanoic acid contained isotopic carbon distributed equally in the carboxyl and carbonyl carbons. The theory of β-oxidation-condensation predicts such an equal distribution of isotope. Acetoacetate produced from carboxyl tagged octanoic acid by a process of multiple alternate oxidation should theoretically contain isotope only in the carboxyl position. The classical successive β-oxidation theory of Knoop would predict the formation of non-isotopic acetoacetate from carboxyl labelled octanoic acid.

Using isotopic carboxyl butyrate as substrate in a similar experiment, Medes, Weinhouse and Floyd (91) found that the acetoacetate produced, likewise contained heavy carbon in the carbonyl and carboxyl positions. In this case, however, the carboxyl carbon atom contained a significantly higher concentration of isotope than the carbonyl carbon. This unequal distribution of isotope in the case of butyric acid cannot be interpreted in terms of the theory of multiple alternate oxidation. By definition, multiple alternate oxidation involves the simultaneous oxidation of alternate carbon atoms with cleavage of a carbon-to-carbon bond at each fourth carbon atom. The incomplete distribution of isotope in this case does demonstrate, however, that the formation of the β-keto acid may precede the carbon-to-carbon cleavage.

Although the results of these authors (143) demonstrate that two-carbon fragments are formed and condense in the process of octanoic acid oxidation, there is still doubt concerning the exact mechanism of this oxidation. It is possible,

for instance, that acetoacetate is formed from octanoic acid by multiple alternate oxidation and that the distribution of the isotopic carbon between the carbonyl and carboxyl carbons results from a cleaving and resynthesis of acetoacetate itself.

The problem of the synthesis of ketone bodies from acetic acid has been studied with carbon isotopes both *in vivo* and *in vitro*. In confirmation of the experiments of MacKay and his co-workers (86), Swenseid, Barnes, Hemingway and Nier (131) have shown that ketone bodies formed by rats after the administration of isotopic acetate contained C^{13} (fig. 1).

Medes, Weinhouse and Floyd have studied the formation of acetoacetic acid from carboxyl labelled acetic acid by a variety of tissues *in vitro*. In liver slices (92), acetic acid is metabolized with an accumulation of acetoacetic acid. The latter compound, when derived from isotopic acetic acid, contained C^{13} in both carboxyl and carbonyl positions. A slight excess was found in the carboxyl position indicating that the β and γ carbon atoms of acetoacetate may have been derived in part from an endogenous source. Kidney and heart slices oxidized acetate without the accumulation of ketone bodies. If non-isotopic acetoacetate was added to kidney slices, however, during the oxidation of isotopic acetate, isotopic acetoacetate could be isolated (144). This was not true of heart tissue (93). Medes, Weinhouse and Floyd have interpreted their data as indicating that two pathways for acetate utilization exist in animal tissues.

From the information so far available, it would appear that, as proposed by MacKay, fatty acids are metabolized by β-oxidation into a two-carbon intermediate which may condense to form acetoacetate. Those two-carbon units which do not condense in this manner may undergo a variety of reactions. They may be oxidized to CO_2, used in the synthesis of cholesterol or fatty acids, or act as acetylating agents. Other isotopic evidence to follow is concerned with the rôle of a two-carbon compound in these reactions and emphasizes the importance of this compound in metabolism.

b. *Sources and reactions of acetate:* As previously mentioned, the early experiment of Schoenheimer and Rittenberg (118) demonstrated that a two-carbon unit is split from stearic acid forming palmitic acid. In a reverse manner, palmitic acid may yield stearic acid (130). The incorporation of deuterium along the entire fatty acid chain by animals synthesizing fat from carbohydrate in the presence of deuterium is further evidence that this synthesis involves the condensation of a number of short chained compounds (115). Recently, Rittenberg and Bloch (104) have shown that isotopic acetic acid may contribute its carbon atoms to the synthesis of fatty acids. Acetic acid labelled in the methyl group with deuterium and in the carboxyl group with heavy carbon was administered to rats and mice for a period of eight days. The fat isolated from their tissues contained both deuterium and heavy carbon in excess of the D_2O and $C^{13}O_2$ concentrations of the fluids. The finding that the C^{13} concentration of the carboxyl carbon atoms of the saturated fatty acids was approximately twice that of the C^{13} concentration in all the carbon atoms of the total molecule suggested that the fatty acids are synthesized by successive condensations of two-carbon units.

The conversion of acetic acid to cholesterol has been studied by Bloch and

Rittenberg (11) in considerable detail. Cholesterol isolated after the feeding of deuterio acetate contained deuterium in both the nucleus and side chain, thus indicating that the synthesis of cholesterol involves the condensation of small carbon units, either acetate itself or a derivative of acetate. In a search for the identity of this cholesterol precursor, Bloch and Rittenberg have fed a variety of deuterio-marked compounds. In no instance have any of the deuterio substituted compounds administered under comparable conditions resulted in higher deuterium concentration in the isolated cholesterol than after feeding deuterio acetate itself. These substances are believed to participate in cholesterol formation by virtue of the fact that they yield acetate during the course of their catabolism.

Although the synthesis of deuterio cholesterol has been used as a measure of the ability of certain deuterium-containing compounds to yield deuterio acetate, Bloch and Rittenberg (12, 10) have recently studied the acetylation of phenyl amino butyric acid in this respect. Their work on this subject was based on the original finding of Bernhard (3) that deuterio acetyl derivatives were excreted when a foreign amine and deuterio acetic acid were administered to dogs. In order to compare the conversion to acetyl of various substances containing different concentrations of deuterium and administered in different doses, Bloch and Rittenberg have compared their results in terms of their coefficients of utilization. The acetylating agent itself should have the highest coefficient of utilization. The administration of deuterium-labelled acetate, ethyl alcohol, alanine, β,γ-dideuterio butyrate, ethyl deuterio myristate, deuterio dl-leucine, deuterio isovalerate, α,β-dideuterio butyrate and deuterio n-valerate yielded deuterium-containing acetyl groups in the order named. The acetyl group of l-acetyl phenyl amino butyric acid excreted after the feeding of deuterio propionate, succinate, valine, isobutyrate, ethyl 10,11 dideuterio undecylate and ethyl 9,10 dideuterio stearate contained no deuterium. The accuracy of the value assigned to the coefficient of utilization depends upon the amount of the administered material catabolized and the extent to which deuterium is removed from the carbon atom through oxidation and labilization reactions. Thus, deuterium on the α carbon of acetoacetic acid is somewhat more labile than that of the ω carbon. Because of this labilization of deuterium on the α carbon atom of a β-keto acid, very little deuterium is found in the acetyl group of the excreted acetyl phenyl amino butyric acid after administration of deuterio fatty acids. However, as expected, deuterium in the ω position of ethyl deuterio myristate appears in the acetyl group.

Since deuterium in the β carbon of alanine is readily converted into the acetyl group, it is assumed that two carbons of pyruvic acid are available for acetylation reactions. The finding that deuterium of α,β-dideuterio propionic acid was not incorporated into the acetyl group of acetyl phenyl amino butyric acid indicates that pyruvic acid is not an intermediate in the oxidation of propionic acid, as has been long supposed. Bloch and Rittenberg believe instead that oxidation at the β carbon atom is the first step in the catabolism of propionic acid.

The extent to which these deuterium-containing compounds were converted to cholesterol very closely paralleled their ability to acetylate phenyl amino butyric acid (12).

Evidence obtained by the isotopic method has thus given support to the belief that fatty acids are metabolized with the formation of a two-carbon intermediate. Other experiments reported by Muñoz and Leloir (96, 80) and by Lehninger (78) indicate that phosphorylation reactions also play a rôle in fatty acid oxidation. The components of the enzyme system oxidizing fatty acids in liver extracts have been described by the former investigators. The addition of cytochrome c, fumaric acid, inorganic phosphate and adenylic acid to the extract is essential for the optimal oxidation of the fatty acids to ketone bodies.

Lehninger (78) has studied the factors involved in the oxidation of octanoic acid by rat liver homogenates and has confirmed the results of Muñoz and Leloir. Under the conditions of these experiments, adenosine triphosphate and adenosine diphosphate could be substituted for fumaric acid, the oxidation of which is presumably responsible for the phosphorylation of adenylic acid. From these experiments, Lehninger has suggested the intermediate formation of acyl phosphates in octanoic acid oxidation. Preliminary experiments indicate that synthetic acyl phosphates having from 8 to 16 carbon atoms do not require adenosine triphosphate for oxidation. After hydrolytic destruction by acid, these synthetic preparations do require adenosine triphosphate activation. These results coupled with the isotopic experiments of Weinhouse et al. indicate that phosphorylation and the formation of a two-carbon intermediate are important stages in the oxidation of fatty acids.

c. *Complete oxidation of fatty acids and ketone bodies.* The reactions concerned with the complete oxidation of fatty acids and ketone bodies to carbon dioxide and water have for many years remained obscure. The possibility of the conversion of fatty acids and their derivatives into carbohydrate during their metabolism has undergone critical experimental examination both *in vitro* and *in vivo*. Weil-Malherbe (142), for example, has reported that there is an increase of the carbohydrate content of rat liver slices metabolizing acetoacetate. Stadie, Zapp and Lukens (124), however, accounted for this newly formed carbohydrate on the basis of the disappearance of carbohydrate precursors preformed in the tissue. At present, there are no decisive experiments *in vivo* demonstrating the conversion of the naturally occurring fatty acids to carbohydrate. Even such a readily metabolized fatty acid as butyric acid does not give rise to liver glycogen in the normal rat (43, 40, 39, 42) nor to "extra carbohydrate" in the phlorhizinized animal (103). Although the application of classical analytical methods has given little or no support to the theory of conversion of fatty acids to carbohydrates, it has not been possible to deny that fatty acids may be metabolized by way of reactions common to carbohydrate oxidation even though "new carbohydrate" is not formed.

The report by Lynen (85) and Virtanen and Sundman (140) that acetate is metabolized by way of the tricarboxylic acid cycle in yeast has been an impetus for further investigation of the metabolism of this compound in mammalian

tissues. Breusch (20, 21) and Wieland and Rosenthal (146) have independently reported the existence of an enzyme system in mammalian tissues which catalyzes the condensation of oxalacetate and acetoacetate or acetate forming citrate. Using kidney breis, Wieland and Rosenthal found that when acetate, acetoacetate, or oxalacetate was incubated separately, small amounts of citrate were formed. The yield was considerably increased, however, by incubating oxalacetate with either of the other two substances. Using muscle extracts, Breusch has obtained essentially the same results. Moreover, Breusch has reported that his muscle extracts are not specific for the oxidation of acetoacetate alone, but can oxidize β keto derivatives of long chained fatty acids.

Krebs and Eggleston (75), however, have recently criticized the work of these investigators. Using minced sheep heart, they confirmed the stimulatory effect of dicarboxylic acids on the disappearance of acetoacetate, and the formation of citric acid. This effect was greater under anaerobic than under aerobic conditions. A quantitative determination of the products of reaction of an anaerobic experiment showed that the extra acetoacetate removed by interaction with oxalacetate could be accounted for as β hydroxybutyric acid. Krebs and Eggleston have explained their results on the basis of the following equations:

$$\left.\begin{array}{l}\text{3 oxalacetic acid} \rightarrow \text{malic acid} + \text{tricarboxylic acid} + 2\ CO_2 \\ \text{malic acid} + \text{acetoacetic acid} \rightarrow \text{oxalacetic acid} + \beta\ \text{hydroxybutyric acid}\end{array}\right\} (17)$$

According to these authors, acetoacetic acid acts as a hydrogen acceptor converting malic acid into oxalacetic acid and hence makes more of this latter compound available for tricarboxylic acid synthesis. Under the conditions of their experiments, oxalacetate did not stimulate acetoacetate disappearance aerobically.

The problem of the oxidative conversion of acetate and acetoacetate into the acids of the tricarboxylic acid cycle has also recently been studied using C^{13} (26). Citrate, α-ketoglutarate, succinate, fumarate, malate, or oxalacetate, when added in proper concentrations to homogenates of guinea pig kidney cortex, stimulates aerobically the disappearance of acetoacetate. Fumaric and α-ketoglutaric acids were isolated after adding isotopic acetoacetate together with an excess of some dicarboxylic acid of the tricarboxylic acid cycle. A similar study was made of the metabolism of acetate *in vitro* (26). All of the isolated acids contained an excess of C^{13} in an order of magnitude which would be expected if the acids of the tricarboxylic acid cycle are a major pathway of acetate and acetoacetate oxidation. These results are in disagreement with the interpretations that Krebs and Eggleston made of their experiments.

The fumaric acid isolated after the metabolism of isotopic acetoacetate contained isotopic carbon only in the carboxyl positions. The α-ketoglutarate contained an excess of C^{13} in both carboxyl groups, but the C^{13} in the carboxyl group γ to the keto group was considerably greater than in the other carboxyl position. If citric acid were an intermediate in acetoacetate oxidation, as postulated by Breusch and by Wieland and Rosenthal, the two carboxyl carbon atoms of α-ketoglutarate should have contained equal concentrations of C^{13}.

The oxidation of acetoacetate may occur by a mechanism similar to that suggested by Wood et al. (154) for pyruvate. In this series of reactions (fig. 1), cisaconitic acid is the condensation product of oxalacetic acid and a two-carbon compound, either acetic acid or a derivative of acetic acid.

The finding that acetic and acetoacetic acids may be metabolized by way of a tricarboxylic acid cycle leads to modification of the present theory of the oxidation of pyruvic acid. It has been previously proposed that oxalocitraconic acid is the product of condensation of pyruvic and oxalacetic acids and that this condensation product compound is further oxidized to cisaconitic acid (71). An alternative interpretation is that pyruvate is first oxidatively decarboxylated to a two-carbon intermediate before being condensed with oxalacetate to form cisaconitate. By such a series of reactions, pyruvate, acetoacetate and fatty acids would have a common intermediate of metabolism (fig. 1). There is objection to regarding acetate as an intermediate of pyruvate oxidation since the oxidative metabolism of pyruvate is considerably greater than acetate in several tissues (71). It has often been suggested that a more metabolically active form of acetate is the product of pyruvate oxidation rather than acetate itself (82, 67, 90). Of these possible active two-carbon intermediates, acetyl phosphate has received greatest attention (82). Knowledge, however, of the reactions of acetyl phosphate in mammalian tissues is so limited that it is impossible to assign it a rôle in oxidative reactions of mammalian tissues at this time.

The mechanism of oxidation of short chain fatty acids has been studied *in vivo* with isotopic carbon (25). Acetic, propionic or butyric acid, each labelled with C^{11} in the carboxyl position, was administered to fasted rats together with glucose to insure an adequate formation of glycogen during a two hour experimental period. The experiments were conducted on the hypothesis that if any of these fatty acids were oxidized via pathways which involved any of the carbohydrate intermediates or any compounds with which the carbohydrates were in dynamic equilibrium, the isotopic carbon of the fatty acid fed should be found in newly formed glycogen carbon. The amount of radioactivity present in the glycogen carbon should be directly proportional to the fraction of the total isotopic material oxidized via the carbohydrate pathway and also directly proportional to the ratio $\frac{\text{amount of glycogen deposited}}{\text{size of effective carbohydrate pool}}$. Thus, the greater the proportion of the carbohydrate pool converted to glycogen, the greater the radioactivity in the glycogen.

The glycogen isolated after feeding each of the three fatty acids with glucose contained radioactive carbon. Before drawing conclusions regarding the origin of the glycogen carbon, it was necessary to estimate in each case the amount of radioactivity which resulted from the incorporation of the $C^{11}O_2$ produced during the metabolism of the acids. All of the radioactivity of the glycogen carbon after acetate feeding could be accounted for by the incorporation of $C^{11}O_2$ derived from the metabolism of the labelled acetate. The glycogen isolated after feeding isotopic butyrate or propionate contained an amount of isotopic carbon significantly in excess of that originating from isotopic CO_2. These results indi-

cated that two or more carbon atoms of the butyric and propionic acid molecules contributed in part to the glycogen carbon, but the results of the experiments with acetate gave no indication of two-carbon unit incorporation in glycogen.

Rittenberg and Bloch (104) have repeated these experiments under somewhat different conditions. They included acetate labelled with C^{13} in the carboxyl position in the diet of white rats for a week. The liver glycogen isolated at the end of the experiment contained a small excess of C^{13} which could not be entirely accounted for by carbon dioxide assimilation.

The question of the conversion of acetic acid into glycogen has also been studied by Lifson, Lorber and Wood (81). They have isolated liver glycogen from rats two hours after receiving glucose and carboxyl isotopic acetate, and degraded it by chemical and bacteriological methods for the purpose of identifying the position of the isotope in the glucose unit of the glycogen. C^{13} was found exclusively in the third and fourth carbon atoms of the glucose unit. These positions likewise contained isotope when glycogen was formed from glucose in the presence of isotopic bicarbonate (147). The similarity of the location of C^{13} in the glucose molecule after the administration of carboxyl isotopic acetate and isotopic bicarbonate would have been expected in view of the fact that during the metabolism of both of these compounds, carboxyl isotopic organic acids of the tricarboxylic acid cycle may be formed (fig. 2). Lifson and his co-workers are repeating these experiments with acetic acid labelled on both carbon atoms (81). The finding of C^{13} in the glucose molecule other than the third and fourth positions would definitely establish whether or not acetate carbon enters directly as a two-carbon unit into the glycogen molecule. The experiments of Lifson et al. with carboxyl isotopic acetate have indicated that the reaction

$$CH_3C^{13}OOH + CO_2 + 2H \rightarrow CH_3C^{13}OCOOH + H_2O \qquad (18)$$

does not take place to a significant extent in the rat. Glucose units derived from pyruvate formed according to this reaction would have contained C^{13} in the first or second and fifth or sixth carbon atoms (fig. 3).

Further evidence in vivo for the conversion of acetic acid into components of the tricarboxylic acid cycle has been supplied by experiments of Rittenberg and Bloch (105). After feeding carboxyl isotopic acetic acid to rats, aspartic and glutamic acids containing isotope in both carboxyl positions have been isolated.

It is not necessary to postulate the over-all conversion of fatty acids to carbohydrate to explain the appearance of isotopic carbon in glycogen isolated from rats fed isotopic fatty acids and non-isotopic glucose. If butyric acid, for example, is oxidized by the series of reactions proposed for acetoacetic acid, oxalacetic acid containing isotopic carboxyl carbon would be formed according to the following reaction:

$$CH_3 \cdot CH_2 \cdot CH \cdot C^{11}OOH + 2O \rightarrow CH_3 \cdot C^{11}OOH + CH_3 \cdot COOH \qquad (19)$$

$$CH_3 \cdot C^{11}OOH + HOOC \cdot CH_2 \cdot CO \cdot COOH \xrightarrow{\text{several reactions}} 2\,CO_2 + HOOC \cdot CH_2 \cdot CO \cdot C^{11}OOH \qquad (20)$$

By these reactions, however, there is no increase in the total amount of carbohydrate precursors since one mole of non-isotopic oxalacetate is used up for every mole of isotopic oxalacetate formed. In the experiments of Buchanan, Hastings and Nesbett (25), isotopic carbon is presumed to enter the glycogen molecule synthesized from administered glucose merely because of the reactions which rapidly bring glucose and the isotopic dicarboxylic acids formed during butyrate oxidation into equilibrium with each other.

On the basis of the above theory of fatty acid oxidation, it is necessary to distinguish between two types of "carbohydrate" forming substances. The first group consists of immediate carbohydrate precursors (e.g., lactic acid) and substances (e.g., propionic acid) which are readily converted into these precursors. These materials upon being fed in sufficient quantity eventually result in an increase in the carbohydrate content of the animal's body and consequently cause glycogen formation to take place. The second group, represented by butyrate or acetoacetate, may be converted to carbohydrate precursors through a condensation reaction with oxalacetic acid. An over-all increase in carbohydrates does not result, however, by this reaction. It would seem, therefore, that the appearance of butyrate carbon in glycogen does not constitute the conversion of fat to carbohydrate in the usually accepted sense.

4. *Metabolism of amino acids.* a. *Rôle of amino acids in carbohydrate metabolism.* Experiments with labelled amino acids have provided additional information on the nature of the carbohydrate metabolic pool. Gurin and Wilson (56) have studied the metabolism of alanine labelled in the carboxyl position with C^{13} in the phlorhizinized rat. They found that the "extra sugar" excreted contained as little as 2.5 per cent of the injected alanine. Although measurement was not made of the excretion of carbon dioxide during the metabolism of alanine, undoubtedly a portion of the alanine carboxyl carbon must have appeared in the respiratory carbon dioxide. It is true that a portion of this expired carbon isotope could have arisen by labilization reactions similar to those described for carboxyl-tagged lactic acid as described by Conant et al. (33).

The problem of the conversion of glycine to glycogen by the rat has been studied by Olsen, Hemingway and Nier (101). Using carboxyl-tagged glycine, they have found that only one of every 29 carbons of newly formed glycogen had its origin from the administered isotopic glycine carbon. Only a small amount of this could have arisen from CO_2 incorporation derived from metabolized glycine.

The conversion of glycine to glycogen is of singular interest because this reaction does not take place until six hours after the administration of glycine (88). The metabolism of glycine starts immediately upon administration, however, since isotopic carbon dioxide was readily excreted in the respiratory gases. Only sixty per cent of the total amount administered was accounted for.

Part of the remainder presumably contributed to the metabolic pool of glycogen precursors. The reason for the delay in glycogen formation and the details of the reactions of glycine subsequent to its administration require further investigation.

b. *Biological synthesis of cystine.* The synthesis of cystine in the rat has been

studied with isotopes of sulfur and carbon. Tarver and Schmidt (133) using radioactive S^{35} methionine demonstrated that the sulfur of methionine is converted into cystine sulfur. The metabolic origin of the carbon moiety of cystine, however, has been in doubt. In order to ascertain whether methionine contributes its carbon chain to the synthesis of cystine, du Vigneaud and co-workers have synthesized (65) and administered (139) methionine labelled with S^{34} and with C^{13} in the β and γ positions. Cystine isolated from the hair of these rats contained large amounts of S^{34} but no C^{13}, thus demonstrating that the carbon of cystine is not derived from methionine. These results are in accord with previous data obtained by du Vigneaud and co-workers and by Stetten. Binkley and du Vigneaud (8) have reported that liver slices synthesize cysteine in the presence of both homocysteine and serine. Methionine could not be substituted for homocysteine in this isolated system, however. They have postulated that cysteine is formed by the splitting of cystathionine (138, 7, 6), a thio ether, resulting from the interaction of serine and homocysteine. Stetten (126), independently, found that rats fed serine containing N^{15} form cystine with relatively high concentrations of N^{15}. Since the N^{15} of the cystine was greater than that of glutamic acid, it could not be attributed to transamination reactions, but rather to the direct utilization of serine in cystine synthesis.

The relatively few experiments so far carried out with amino acids labelled with isotopic carbon illustrate the type of valuable information which can be obtained and are doubtless but forerunners of more extensive studies along similar lines.

SUMMARY

From the above review one may summarize certain outstanding additions to knowledge which have modified the points of view from which one must approach problems of metabolism.

a. CO_2 may no longer be considered as simply an end-product of metabolism but must be thought of as an essential compound in certain metabolic reactions in higher as well as lower forms of life.

b. Instead of considering compounds, such as lactic acid, as specific precursors of more complex compounds, such as glycogen, they must be considered as contributors to the metabolic pool from which the complex compounds are formed.

c. More intimate information is now available of the reactions relating carbohydrate, fat and protein precursors and their contribution to the metabolic pool of each.

This is but the beginning. It is fully to be expected that when these tools which permit the tracing of atoms in the body have widespread use, it will be found that many of the earlier conclusions regarding the fate of substances will require drastic revision, and that many problems which hitherto had no means of approach will become vulnerable to attack.

The authors are indebted to many colleagues who have made helpful suggestions during the preparation of this review, and particularly to Prof. D. Wright Wilson, who was good enough to endure and improve several revisions.

REFERENCES

References primarily concerned with isotopes are marked *.

(1)*Barker, H. A., S. Ruben and M. D. Kamen. Proc. Natl. Acad. Sci. **26:** 426, 1940.
(2) Benedict, S. R. and E. Osterberg. J. Biol. Chem. **56:** 229, 1923.
(3)*Bernhard, K. Ztschr. physiol. Chem. **267:** 91, 1940; **267:** 99, 1940.
(4)*Bernhard, K. and R. Schoenheimer. J. Biol. Chem. **133:** 707, 1940.
(5)*Bernhard, K. and R. Schoenheimer. J. Biol. Chem **133:** 713, 1940.
(6) Binkley, F. J. Biol. Chem. **155:** 39, 1945.
(7) Binkley, F., W. P. Anslow, Jr. and V. du Vigneaud. J. Biol. Chem. **143:** 559, 1942.
(8) Binkley, F. and V. du Vigneaud. J. Biol. Chem. **144:** 507, 1942.
(9) Blixenkrone-Møller, N. Ztschr. physiol. Chem. **252:** 117, 1938; **252:** 137, 1938.
(10)*Bloch, K. J. Biol. Chem. **155:** 255, 1944.
(11)*Bloch, K. and D. Rittenberg. J. Biol. Chem. **145:** 625, 1942.
(12)*Bloch, K. and D. Rittenberg. J. Biol. Chem. **155:** 243, 1944.
(13)*Bloch, K. and R. Schoenheimer. J. Biol. Chem. **131:** 111, 1939.
(14)*Bloch, K. and R. Schoenheimer. J. Biol. Chem. **138:** 167, 1941.
(15)*Bloch, K., R. Schoenheimer and D. Rittenberg. J. Biol. Chem. **138:** 155, 1941.
(16) Borsook, H. and G. L. Keighley. Proc. Royal Soc. London, Series B **118:** 488, 1935.
(17)*Boxer, G. E. and DeW. Stetten, Jr. J. Biol. Chem. **153:** 607, 1944.
(18)*Boxer, G. E. and DeW. Stetten, Jr. J. Biol. Chem. **155:** 237, 1944.
(19) Braunstein, A. E. and M. G. Kritzmann. Enzymologia **2:** 129, 1937.
(20) Breusch, F. L. Science **92:** 490, 1943.
(21) Breusch, F. L. Compt. rend. ann. arch. Soc. Turque sc. phys. Nat. **10:** 55, 1942–43.
(22) Brown, R. W., H. G. Wood and C. H. Werkman. Arch. Biochem. **5:** 423, 1944.
(23)*Buchanan, J. M. Thesis, Glycogen formation, Harvard University, 1943.
(24)*Buchanan, J. M., A. B. Hastings and F. B. Nesbett. J. Biol. Chem. **145:** 715, 1942.
(25)*Buchanan, J. M., A. B. Hastings and F. B. Nesbett. J. Biol. Chem. **150:** 413, 1943.
(26)*Buchanan, J. M., W. Sakami, S. Gurin and D. W. Wilson. J. Biol. Chem. **157:** 747, 1945.
(27) Butts, J. S., C. H. Cutler, L. Hallman and H. J. Deuel, Jr. J. Biol. Chem. **109:** 597, 1935.
(28)*Carson, S. F., J. W. Foster, S. Ruben and H. A. Barker. Proc. Natl. Acad. Sci. **27:** 229, 1941.
(29)*Carson, S. F., J. W. Foster, S. Ruben and M. D. Kamen. Science **92:** 433, 1940.
(30)*Carson, S. F. and S. Ruben. Proc. Natl. Acad. Sci. **26:** 422, 1940.
(31)*Carson, S. F., S. Ruben, M. D. Kamen and J. W. Foster. Proc. Natl. Acad. Sci. **27:** 475, 1941.
(32)*Clarke, H. T. in R. Schoenheimer. Dynamic state of body constituents, Cambridge, 1942.
(33)*Conant, J. B., R. D. Cramer, A. B. Hastings, F. W. Klemperer, A. K. Solomon and B. Vennesland. J. Biol. Chem. **137:** 557, 1941.
(34) Cori, C. F. and G. T. Cori. J. Biol. Chem. **81:** 389, 1929.
(35) Cori, C. F. Biological Symposia **5:** 131, 1941.
(36) Cori, C. F. and G. T. Cori. J. Biol. Chem. **79:** 309, 1928.
(37)*Cramer, R. D. and G. B. Kistiakowsky. J. Biol. Chem. **137:** 549, 1941.
(38) Dakin, H. D. J. Biol. Chem. **6:** 221, 1909.
(39) Deuel, H. J., Jr., J. S. Butts, H. Blunden, C. H. Cutler and L. Knott. J. Biol. Chem. **117:** 119, 1937.
(40) Deuel, H. J., Jr., J. S. Butts, L. F. Hallman and C. H. Cutler. J. Biol. Chem. **112:** 15, 1935–36.
(41) Deuel, H. J., Jr., L. F. Hallman, J. S. Butts and S. Murray. J. Biol. Chem. **116:** 621, 1936.
(42) Deuel, H. J., Jr., C. Johnston, M. G. Morehouse, H. S. Rollman and R. J. Winzler. J. Biol. Chem. **157:** 135, 1945.

(43) Eckstein, H. C. J. Biol. Chem. **102:** 591, 1933.
(44) von Euler, H., E. Adler, G. Günther and N. B. Das. Ztschr. physiol. Chem. **254:** 61, 1938.
(45) Evans, E. A., Jr. Biochem. J. **34:** 829, 1940.
(46)*Evans, E. A., Jr. A symposium on respiratory enzymes, 197, Madison, 1942.
(47)*Evans, E. A., Jr. and L. Slotin. J. Biol. Chem. **136:** 301, 1940; **141:** 439, 1941.
(48)*Evans, E. A., Jr. and L. Slotin. J. Biol. Chem. **136:** 805, 1940.
(49)*Evans, E. A., Jr., L. Slotin and B. Vennesland. J. Biol. Chem. **143:** 565, 1942.
(50)*Evans, E. A., Jr., B. Vennesland and L. Slotin. J. Biol. Chem. **147:** 771, 1943.
(51) Folin, O. Am. J. Physiol. **13:** 117, 1905.
(52) Folin, O. In Festschrift für O. Hammarsten, Upsala, iii, 1, 1906.
(53)*Foster, J. W., S. F. Carson, S. Ruben and M. D. Kamen. Proc. Natl. Acad. Sci. **27:** 590, 1940.
(54) Franck, J. and H. Gaffron. Advances in enzymology **1:** 199, New York, 1941.
(55) Gaffron, H. Biol. Rev. **19:** 1, 1944.
(56)*Gurin, S. and D. W. Wilson. Fed. Proc. **1:** 114, 1942.
(57)*Heidelberger, M., H. P. Treffers, R. Schoenheimer, S. Ratner and D. Rittenberg. J. Biol. Chem. **144:** 555, 1942.
(58) Hurtley, W. H. Quart. J. Med. **9:** 301, 1916.
(59) Jowett, M. and J. H. Quastel. Biochem. J. **29:** 2159, 1935.
(60) Kalckar, H. Chem. Rev. **28:** 71, 1941.
(61)*Kalnitsky, G. and C. H. Werkman. J. Bact. **44:** 256, 1942.
(62) Kalnitsky, G. and C. H. Werkman. Arch. Biochem. **2:** 113, 1943.
(63) Kalnitsky, G. and C. H. Werkman. Arch. Biochem. **4:** 25, 1944.
(64)*Kalnitsky, G., H. G. Wood and C. H. Werkman. Arch. Biochem. **2:** 269, 1943.
(65)*Kilmer, G. W. and V. du Vigneaud. J. Biol. Chem. **154:** 247, 1944.
(66) Knoop, F. Beitr. chem. Physiol. u. Path. **6:** 150, 1905.
(67) Knoop, F. and E. Kertess. Ztschr. physiol. Chem. **71:** 252, 1911.
(68) Krampitz, L. O. and C. H. Werkman. Biochem. J. **35:** 595, 1941.
(69)*Krampitz, L. O., H. G. Wood and C. H. Werkman. J. Biol. Chem. **147:** 243, 1943.
(70) Krebs, H. A. Ann. Rev. Biochem. **12:** 529, 1943.
(71) Krebs, H. A. Advances in enzymology **3:** 191, New York, 1943.
(72) Krebs, H. A. and L. V. Eggleston. Biochem. J. **34:** 442, 1940.
(73) Krebs, H. A. and L. V. Eggleston. Biochem. J. **34:** 1383, 1940.
(74) Krebs, H. A. and L. V. Eggleston. Biochem. J. **35:** 676, 1941.
(75) Krebs, H. A. and L. V. Eggleston. Nature **154:** 209, 1944.
(76) Krebs, H. A. and W. A. Johnson. Enzymologia **4:** 148, 1937.
(77) Lardy, H. A. and J. A. Ziegler. J. Biol. Chem. **159:** 343, 1945.
(78) Lehninger, A. L. J. Biol. Chem. **154:** 309, 1944; **157:** 363, 1945.
(79) Leloir, L. F. and J. M. Muñoz. Biochem. J. **33:** 734, 1944.
(80) Leloir, L. F. and J. M. Muñoz. J. Biol. Chem. **153:** 53, 1944.
(81)*Lifson, N., V. Lorber and H. G. Wood. Fed. Proc. **4:** 47, 1945.
(82) Lipmann, F. Advances in Enzymology **1:** 99, New York, 1941.
(83) Lipmann, F. and L. C. Tuttle. J. Biol. Chem. **154:** 725, 1944; **158:** 505, 1945.
(84)*Lorber, V., A. Hemingway and A. O. Nier. J. Biol. Chem. **151:** 647, 1943.
(85) Lynen, F. Ann. **552:** 270, 1942; **554:** 40, 1945.
(86) MacKay, E. M., R. H. Barnes, H. O. Carne and A. N. Wick. J. Biol. Chem. **135:** 157, 1940.
(87) MacKay, E. M., A. N. Wick and C. P. Barnum. J. Biol. Chem. **136:** 503, 1940.
(88) MacKay, E. M., A. N. Wick and H. O. Carne. J. Biol. Chem. **132:** 613, 1940.
(89) Madden, S. C. and G. H. Whipple. This Journal **20:** 194, 1940.
(90) Martius, C. Ztschr. physiol. Chem. **279:** 96, 1943.
(91)*Medes, G., S. Weinhouse and N. F. Floyd. J. Biol. Chem. **157:** 35, 1945.
(92)*Medes, G., S. Weinhouse and N. F. Floyd. J. Biol. Chem. **157:** 751, 1945.

(93)*Medes, G., S. Weinhouse and N. F. Floyd. Fed. Proc. **4:** 98, 1945.
(94)*Meyerhof, O., P. Ohlmeyer, W. Gentner and H. Maier-Leibnitz. Biochem. Ztschr. **298:** 396, 1938.
(95) Meyers, V. C. and M. S. Fine. J. Biol. Chem. **14:** 9, 1913.
(96) Muñoz, J. M. and L. F. Leloir. J. Biol. Chem. **147:** 355, 1943.
(97) van Niel, C. B. Advances in enzymology **1:** 263, 1941.
(98) van Niel, C. B. This Journal **23:** 338, 1943.
(99)*van Niel, C. B., J. O. Thomas, S. Ruben and M. D. Kamen. Proc. Natl. Acad. Sci. **28:** 157, 1942.
(100)*Nishina, Y., S. Endo and H. Nakayama. Sci. Papers, Inst. Phys. Chem. Research, Tokyo **38:** 341, 1941.
(100a) Ochoa, S. and E. Weisz-Tabori. J. Biol. Chem. **159:** 245, 1945.
(100b) Ochoa, S. J. Biol. Chem. **159:** 243, 1945.
(101)*Olsen, N. S., A. Hemingway and A. O. Nier. J. Biol. Chem. **148:** 611, 1943.
(102) Pflüger, E. Pflüger's Arch. **54:** 334, 1893.
(103) Ringer, A. I. J. Biol. Chem. **14:** 43, 1913.
(104)*Rittenberg, D. and K. Bloch. J. Biol. Chem. **154:** 311, 1944.
(105)*Rittenberg, D. and K. Bloch. J. Biol. Chem. **157:** 749, 1945.
(106)*Rittenberg, D. and R. Schoenheimer. J. Biol. Chem. **121:** 248, 1937.
(107)*Rittenberg, D. and H. Waelsch. J. Biol. Chem. **136:** 799, 1940.
(108) Rose, W. C. and F. W. Dimmitt. J. Biol. Chem. **26:** 345, 1916.
(109)*Ruben, S. J. Am. Chem. Soc. **65:** 279, 1943.
(110)*Ruben, S., W. Z. Hassid and M. D. Kamen. J. Am. Chem. Soc. **61:** 661, 1939.
(111)*Ruben, S. and M. D. Kamen. Proc. Natl. Acad. Sci. **26:** 418, 1940.
(112)*Ruben, S. and M. D. Kamen. J. Am. Chem. Soc. **62:** 3451, 1940.
(113)*Ruben, S., M. D. Kamen and W. Z. Hassid. J. Am. Chem. Soc. **62:** 3443, 1940.
(114)*Ruben, S., M. D. Kamen and L. H. Perry. J. Am. Chem. Soc. **62:** 3450, 1940.
(115)*Schoenheimer, R. Dynamic state of body constituents. Cambridge, 1942.
(116)*Schoenheimer, R., S. Ratner, D. Rittenberg and M. Heidelberger. J. Biol. Chem. **144:** 541, 1942.
(117)*Schoenheimer, R., S. Ratner, D. Rittenberg and M. Heidelberger. J. Biol. Chem. **144:** 545, 1942.
(118)*Schoenheimer, R. and D. Rittenberg. J. Biol. Chem. **120:** 155, 1937.
(119*)Slade, H. D. and C. H. Werkman. Arch. Biochem. **2:** 97, 1943.
(120)*Slade, H. D., H. G. Wood, A. O. Nier, A. Hemingway and C. H. Werkman. J. Biol. Chem. **143:** 133, 1942.
(121)*Solomon, A. K., B. Vennesland, F. W. Klemperer, J. M. Buchanan and A. B. Hastings. J. Biol. Chem. **140:** 171, 1941.
(122) Stadie, W. C. J. Clin. Investigation **19:** 842, 1941.
(123) Stadie, W. C., J. A. Zapp, Jr. and F. D. W. Lukens. J. Biol. Chem. **132:** 423, 1940.
(124) Stadie, W. C., J. A. Zapp, Jr. and F. D. W. Lukens. J. Biol. Chem. **137:** 63, 1941.
(125) Stare, F. J. and C. A. Baumann. Proc. Roy. Soc., B **121:** 338, 1936.
(126)*Stetten, DeW., Jr. J. Biol. Chem. **144:** 501, 1942.
(127)*Stetten, DeW., Jr. and G. E. Boxer. J. Biol. Chem. **155:** 231, 1944.
(128)*Stetten, DeW., Jr. and G. E. Boxer. J. Biol. Chem. **156:** 271, 1944.
(129)*Stetten, DeW., Jr., G. E. Boxer and B. Klein. Fed. Proc. **4:** 106, 1945.
(130)*Stetten, DeW., Jr. and R. Schoenheimer. J. Biol. Chem. **133:** 329, 1940.
(131)*Swendseid, M. E., R. H. Barnes, A. Hemingway and A. O. Nier. J. Biol. Chem. **142:** 47, 1942.
(132) Szent-Györgyi, A. Ztschr. physiol. Chem. **235:** 1, 1935; **244:** 105, 1936.
(133)*Tarver, H. and C. L. A. Schmidt. J. Biol. Chem. **130:** 67, 1939.
(134)*Ussing, H. H. Skand. Arch. Physiol. **77:** 85, 1937.
(135)*Utter, M. F., F. Lipmann and C. H. Werkman. J. Biol. Chem. **154:** 723, 1944; **158:** 521, 1945.

(136)*Vennesland, B., A. K. Solomon, J. M. Buchanan, R. D. Cramer and A. B. Hastings. J. Biol. Chem. **142:** 371, 1942.
(137)*Vennesland, B., A. K. Solomon, J. M. Buchanan and A. B. Hastings. J. Biol. Chem. **142:** 379, 1942.
(138) du Vigneaud, V., G. B. Brown and J. P. Chandler. J. Biol. Chem. **143:** 59, 1942.
(139)*du Vigneaud, V., G. W. Kilmer, J. R. Rachele and M. Cohn. J. Biol. Chem. **155:** 645, 1944.
(140) Virtanen, A. J. and S. Sundman. Biochem. Ztschr. **313:** 236, 1942.
(141) Voit, C. Cited by O. Folin. Am. J. Physiol. **13:** 117, 1905.
(142) Weil-Malherbe, H. Biochem. J. **32:** 2276, 1938.
(143)*Weinhouse, S., G. Medes and N. F. Floyd. J. Biol. Chem. **153:** 689, 1944; **155:** 143, 1944.
(144)*Weinhouse, S., G. Medes and N. F. Floyd. J. Biol. Chem. **158:** 411, 1945.
(145)*Werkman, C. H. and H. G. Wood. Advances in enzymology **2:** 135, New York, 1942.
(146) Wieland, H. and C. Rosenthal. Ann. Chem. **554:** 241, 1945.
(147)*Wood, H. G., N. Lifson and V. Lorber. J. Biol. Chem. **159:** 475, 1945.
(147a)*Wood, H. G., B. Vennesland and E. A. Evans, Jr. J. Biol. Chem. **159:** 153, 1945.
(148) Wood, H. G. and C. H. Werkman. Biochem. J. **30:** 48, 1936.
(149) Wood, H. G. and C. H. Werkman. Biochem. J. **32:** 1262, 1938.
(150)*Wood, H. G., C. H. Werkman, A. Hemingway and A. O. Nier. J. Biol. Chem. **135:** 789, 1940.
(151)*Wood, H. G., C. H. Werkman, A. Hemingway and A. O. Nier. J. Biol. Chem. **139:** 365, 1941.
(152)*Wood, H. G., C. H. Werkman, A. Hemingway and A. O. Nier. J. Biol. Chem. **139:** 377, 1941.
(153)*Wood, H. G., C. H. Werkman, A. Hemingway and A. O. Nier. Proc. Soc. Exper. Biol. and Med. **46:** 313, 1941.
(154)*Wood, H. G., C. H. Werkman, A. Hemingway and A. O. Nier. J. Biol. Chem. **139:** 483, 1941; **142:** 31, 1942.
(155) Woods, D. D. Biochem. J. **30:** 515, 1936.

YAWNING AND ASSOCIATED PHENOMENA

A. PRICE HEUSNER[1]

Department of Psychiatry, Massachusetts General Hospital and the Department of Neurology and Psychiatry, Harvard Medical School

In the course of some studies (1) on the vasomotor activity in human fingers I noted that definite circulatory changes occur during spontaneous yawning. An attempt to correlate these observations with existing data on yawning revealed that the few pertinent studies of this subject have not found their way into British or American publications. Although yawning and its associated phenomena appear to be of minor importance they are not without interest, and it would seem worth while to make readily available the little that is known and the rather more that has been postulated on this subject. Such is the purpose of this paper.

LITERATURE. *Standard sources.* It is well to review the popular conceptions of yawning as set forth in the standard physiologic, psychologic and lay publications before turning to the special literature.

Of the well-known text-books of physiology eight (2, 3, 4, 5, 6, 7, 8, 9) contain no discussion of yawning. One of them (10), in a section devoted to "modified respiratory movements," defines it in the following terms:

> Various psychic or emotional reactions manifest themselves not only by facial expressions, but also by objective changes in respiratory acts. . . . Yawning is associated with a prolonged inspiration during which the mouth is stretched wide open, followed by a short expiration. (p. 386)

The larger text-book of Schäfer (11), in a discussion of "modified respiratory movements," does not go beyond stating:

> Yawning consists of a deep inspiration accompanied by a wide opening of the mouth and of the glottis, often attended by movements of the arms. (Vol. ii; p. 309)

Even the ponderous German hand-books on physiology treat the subject in cursory fashion. Thus in Hermann's *Handbuch der Physiologie* (12), under the heading "special forms of respiratory movements," yawning is defined and dismissed in these terms:

> [yawning is] a psychically or reflexly excited inspiration during which the vocal cords may be set in motion. (B. IV; H. 2; S. 234)

In Nagel's *Handbuch der Physiologie des Menschen* (13) the topic is introduced and concluded with the following remarks in a section devoted to "special forms of respiration":

> The yawn is a deep inspiration carried out with widely opened glottis and usually with opened mouth; it is frequently accompanied by movements of the arms, *etc.* It is caused by certain psychic influences, fatigue, *etc.* (B. I; S. 27)

---

[1] Present Address: Department of Neurosurgery, Boston City Hospital, Boston, Mass.

From the foregoing it is apparent that while none of these authors makes any conjecture as to the physiologic significance of the act, they have all focused their attention upon its respiratory component and have classified it as a respiratory phenomenon. It is hardly necessary to point out that the subject has not been developed beyond or even through the descriptive phase.

The psychologic literature appears to be equally barren. Yawning has been defined (14) as ". . . .a reflex usually induced by bodily fatigue," but critical studies in support of this conception are wanting. In a recent review of the psychologic aspects of the subject Moore (15) points out that nothing about yawning has been established beyond the notation that it appears very early in extra-uterine life.

Articles for lay consumption are more explicit and display less hesitancy in assigning a significance or purpose to the yawn. This is evident in the following excerpt from a recent *New York Times Magazine* (16):

> Boredom, hunger, overeating, drowsiness and bad ventilation have all been blamed for yawns, but the specific stimulus that starts the reflex is still obscure. The power of suggestion has a lot to do with it. Yawns are catching, as everyone knows, and they can even be self-induced. . . . A yawn has all the beneficial effects of a deep breath of fresh air. It is relaxing and refreshing. It tones up the muscles of the mouth, the chest, the back and even the arms. Whether it has any other psychological or physical effects science doesn't say.

The above views are essentially the same as those which Clarkson (17) submitted to the laity in *Hygeia*. They constitute a modern version of the Delsartean concept (18) which we will review in a later paragraph.

It is of interest to compare all of the foregoing descriptions with the dictionary definitions of the word *yawning* and its alleged synonyms, *pandiculation*, *oscitation* and *chasma*. Standard lay (19, 20) and medical (21) dictionaries agree that the word *yawning* is derived from the Old English verb *ganien* which has always denoted an "opening" in the sense of "to gape." The definitions of yawning given by the lay dictionaries (19, 20) are in such complete harmony that only one need be quoted:

> An involuntary act, excited by drowsiness, *etc.*, and consisting of a deep and long inspiration usually following several successive attempts at inspiration, the mouth, fauces, *etc.*, typically being wide open. (20)

Some medical dictionaries (21) conform to this restricted definition, but others have expanded the meaning to include the associated stretching of the limbs. Thus Dorland (22) offers the following definition of yawning:

> Pandiculation; a deep, involuntary inspiration with the open mouth, often accompanied by the act of stretching.

From this it also appears that Dorland regards pandiculation and yawning as synonyms.

Both lay (19, 20) and medical (21, 22) dictionaries agree that *pandiculation* is derived from the Latin verb *pandiculari*, meaning "to stretch oneself." Of these dictionaries only Dorland, as already noted, offers this word as a synonym

for yawning. The others are in agreement on a more narrow definition of which the following is characteristic:

> A stretching and stiffening, esp. of the trunk and extremities, as when fatigued and drowsy, or after waking from sleep. (20)

All of these dictionaries (19, 20, 21, 22) offer the words *chasma* (from the Latin noun *chasma*. . . .a cleft or an abyss) and *oscitation* (from the Latin verb *oscitare* . . . .to open the mouth wide) as synonyms for the noun *yawn*.

From the above it is evident that the authorities, both lay and medical, have been aware of the complexity of the act under consideration and have been uncertain just how much of it to include under the designation "yawn." On semasiological grounds it would appear desirable to preserve a distinction between yawning and pandiculation, and to employ some such expression as "the yawn-stretch act" when the entire complex is intended.

Special sources. (Occurrence). Depending on the breadth of definition adopted, the occurrence of yawning throughout the animal kingdom is more or less widespread. Crämer (23) points out that the familiar gaping movements of fishes, amphibians and reptiles simulate the oral component of yawning in man, but most observers (24, 25, 26, 27) decline to read true yawning into this superficial resemblance. Whether or not the gaping of these lower forms is accompanied by the respiratory component of yawning and a generalized stretching is not known, and it is best to follow Peiper (27) in reserving judgment on the occurrence of yawning in these classes.

The data on birds are little more complete, although the gaping chicken beating its wings on tip-toe is an everyday spectacle. Here a generalized stretching is seen in conjunction with gaping, but the question of an associated respiratory effort has not been answered by critical observation. It is the impression of both Heinroth (25) and Crämer (23) that true yawning is practiced by birds, but Hauptmann (26) opposes this view.

It is generally agreed (23, 24, 25, 26, 27) that among terrestrial mammals true yawning and stretching are to be observed in primates and carnivores. The situation with regard to herbivora is not so clear, although most authors concede that these animals practice stretching. The failure of the herbivora to exhibit the gaping component of the yawn-stretch act constitutes an objection standing in the way of admitting them into the category of yawners, but Dumpert (24) has pointed out that since herbivora are exclusively nose-breathers, even under conditions of stress, it is unnecessary for them to add the oral component to the yawn-stretch act in order to effect its inspiratory phase.

Before leaving the subject of yawning in animals it is important to mention the possible relationship of its gaping component to the "automatic expressions" described by Darwin (28). The gaping movements of fishes, amphibians and reptiles may belong to this group of emotional expressions, and it appears certain that some of the anthropoids employ the mouth-act in question for emotional attitudinization. Finally, as we shall see, there are those who regard the complete yawn-stretch complex as but another example of the automatic expressions like laughing and crying.

In the case of man little can be added to the previously noted lay observations on occurrence. The idea that yawns are "catching" can be traced back to Mayer (29) who believes that in some individuals the threshold for this so-called psychically released yawning is exceptionally low, and that in such persons autosuggestion may constitute an effective stimulus. Crämer (23) and Schrijver (30) hold that yawning may normally be a symptom of hunger. They report that it is so regarded by some of the Continental laity, and Luciani (31) has been cited as supporting this view. It should be noted, however, that yawning was not among the symptoms of the fasting men observed by Carlson (32).

It has been remarked (24, 33) that among infants and children, as among many animals, yawning is more regularly attended by stretching than is the case among adults. Dumpert (24) looks upon the adult's yawn, without pandiculation, as a fractionation of a primitive yawn-stretch reaction, and he postulates that such a dissociation is the outcome of acquired voluntary inhibitions cultivated in accordance with the demands of "good company." As we shall see in a subsequent paragraph, this concept enjoys some support from the fact that man's ability to yawn without stretching can be lost in certain disorders of the central nervous system. There are other observations, however, which suggest that yawning and stretching are independent albeit frequently associated acts, rather than fundamentally conjoined responses. In the first place, infants under 1 year frequently yawn without stretching and it is hardly permissible to regard this as a fractionation of some basic physiologic pattern by voluntary inhibitions. Secondly, regardless of man's age he displays a tendency to yawn (without stretching) just prior to surrendering to sleep, and to stretch (without yawning) upon waking. This too intimates that factors other than voluntary inhibitions are active in determining whether pandiculation and yawning occur together or separately.

Yawning under circumstances other than those already mentioned is probably always morbid.

Special sources. (Clinical.) Addressing clinicians in 1908 Geigel (34) lamented the fact that his teachers had never mentioned yawning and complained that nowhere could he discover observations elucidating its physiologic or pathologic significance. Similar conditions obtain today although morbid yawning is more common than our neglect of it would imply.

Geigel's contribution to the subject was quite limited. He emphasized the value of yawning as a favorable prognostic sign in certain acute illnesses and noted its occurrence in hysterics. Several observers (24, 27, 35, 36) have remarked the frequency of yawning after severe hemorrhage. Peiper (27) seeks to explain this on the basis of his theory (*vide infra*) of lowered excitability of the central nervous system, whereas Nash (36) attributes it to anoxia and/or arterial hypotension. There are no experimental data supporting either explanation.

Crämer (23) and Schrijver (30) have reported the frequency of excessive yawning in diseases of the gastro-intestinal tract, especially that of duodenal ulcer, wherein crises of yawning sometimes parallel exacerbations of symptoms. The association of deranged stools and flatus with morbid yawning so impressed

Crämer that he postulated an auto-intoxication by bacterial toxins arising from within the bowel as the causative agent. He reviews the rather meagre evidence supporting this view.

Neurologists have met with abnormal yawning rather more frequently than others. Oppenheim (37), Geigel (34) and Lewy (35) have encountered it in hysteria, and its occurrence as an aura in epilepsy has been remarked by Oppenheim (37), Lewy (35), Wilson (38) and Penfield (39). In Wilson's experience this aura has been associated with his so-called "peri-ventricular" or "visceral" epilepsy wherein the yawn is followed by voiding of urine, giddiness, headache and sometimes vomiting. He suggests that such cases depend on disorder of visceral centers in the third ventricle regulating sleep and fluid interchange, with subsequent spread to others on the floor of the fourth ventricle. Penfield (p. 110) gives the clinical details of such a case.

Among the manifold sequelae of epidemic encephalitis is that of paroxysmal yawning usually in conjunction with other, though varied, neurologic disturbances. Mayer (29), Wilson (38) and Sicard and Paraf (40) describe cases from their personal experience, but it has not been possible to ascribe the symptom to proven lesions at particular foci in the neuro-axis. In this connection it is of interest to record certain impressions which Yakovlev (41) has formed with regard to derangements of yawning and pandiculation in other disorders involving the basal ganglia. In his experience Huntington's Choreics exhibit a propensity for yawning and more especially for stretching which stands in sharp contrast to the behaviour of Parkinsonics who rarely yawn and never stretch. Yakovlev emphasizes that these impressions have been derived from a study of relatively few cases and stand in need of confirmation by a test on more material.

Excessive yawning is frequently seen in patients harboring expanding intracranial lesions, notably cerebellar abscess or tumor (35, 36, 37, 38, 42). When attended by abnormal drowsiness the sign is without localizing value (37, 38), but its occurrence in an alert patient suggests to some (38) that a thorough search be made for third ventricular and hypothalamic signs. Nash (36) hazards an explanation for the yawning seen in the usual run of brain tumor cases. He states:

> The occurrence of yawning, which is essentially an atypical respiratory act, in cases of brain tumor can be explained by irritation of the brain resulting in "afferent" impulses to the medullary respiratory center, which modify its rhythmic activity. (p. 92)

Special sources. (Investigative.) Mayer (29) alone has attempted a systematic investigation of the yawn-stretch act. By means of the roentgen-screen, laryngeal mirror and palpation of muscles he has analyzed the sequence of events in man when the complete complex, unmodified by inhibition, is allowed to develop. For convenience of description he divides the act into a number of phases:

I. Inspiratory Phase (duration: 4.4 to 6.8 sec.)
  A. Initial Part
  B. Acme (duration: 2.5 sec.)
II. Expiratory Phase

During the initial part of the inspiratory phase one observes a widening of the chest, a descent of the diaphragm and larynx, an elevation of the *ali nasi* and soft palate, a downward and backward displacement of the tongue, an abduction of the vocal cords and an opening of the mouth. This is occasionally attended by a subjective crackling in the ears which is thought (43) to be due to a contraction of the *M. tensor veli palatini* with jerking of its tendon in the bursa between it and the pterygoid hamulus.

When the acme of the complex has been attained one notes that the chest, diaphragm, larynx, tongue, palate and mouth have maintained or accentuated their respective movements of the initial phase. There has, in addition, been a firm closure of the eyes and a rapid development of the stretch component. The number and distribution of muscles recruited for the stretch are subject to variation and may be modified by voluntary inhibition. Neck and arm muscles are most frequently recruited, but those of the legs, trunk and abdomen may participate. The probable influence of such widespread muscular activity upon the venous return, and secondarily upon cardiac filling, rate and output, has been recognized (24, 29), but no studies have been made to confirm this. At the close of this phase saliva occasionally spurts from the submaxillary and sublingual ducts, presumably due to a squeezing of these glands by the contracting oral musculature.

During the expiratory phase one notes a short sharp expiration with gradual relaxation of all previously activated muscles permitting the return of diaphragm, larynx, tongue, *etc.* to their resting positions. Lacrimation and swallowing may occur at this stage. The former is believed (44) to be due to an augmentation of secretion rather than to a simple compression of the lacrymal glands because the tightest possible voluntary blepharospasm cannot squeeze forth tears in an amount sufficient to overflow the conjunctival sac. The swallowing is presumed to depend upon the occurrence of the previously noted salivation. This final phase of the yawn is usually accompanied by some such subjective feeling as relief, satisfaction or refreshment.

Prior to Mayer's (29) fluoroscopic observations, which show the diaphragm to descend and remain relatively fixed in this position from the initiation until the expiratory phase of the yawn, Hauptmann (26) had arrived at a different opinion as to the behaviour of this organ. Using a kymographic method for recording thoracic and abdominal respiration, Hauptmann was led to believe that the rhythmic action of this muscle was but little, if at all, interrupted by yawning. His view must, of course, give way to that suggested by the more refined method of fluoroscopy. It is important to note, however, that Mayer and Hauptmann agree that yawning effects no ultimate increase in aeration of the lungs; they argue that the apneic period following the yawn (see fig. 1 and 2) more than compensates for any temporary increase in aeration afforded by a single deep breath.

The neural structures subserving the complicated yawn-stretch act may conveniently be thought of as comprising an integrative center played upon by diverse afferent elements and discharging over its several efferent channels. Evi-

dence at hand is insufficient, however, to establish the location or even the existence of a "yawn-center" in man or animals. Ellenberger and Scheunert (45) include yawning among the functions of the medulla oblongata in dogs, but they do not submit the evidence supporting this localization. In man a postulated center has been variously located in the *corpus striatum* (23, 35) and "subcortical ganglia" (24, 27, 29) on very unsatisfactory evidence. There are, however, two significant papers bearing on this question. Gamper (46) has seen the complete yawn-stretch act in what he describes as a mid-brain preparation (*Arhinencephalie mit Encephalocele*), and Catel and Krauspe (47) have observed it in a case where only the medulla oblongata and more caudal portions of the nervous system were

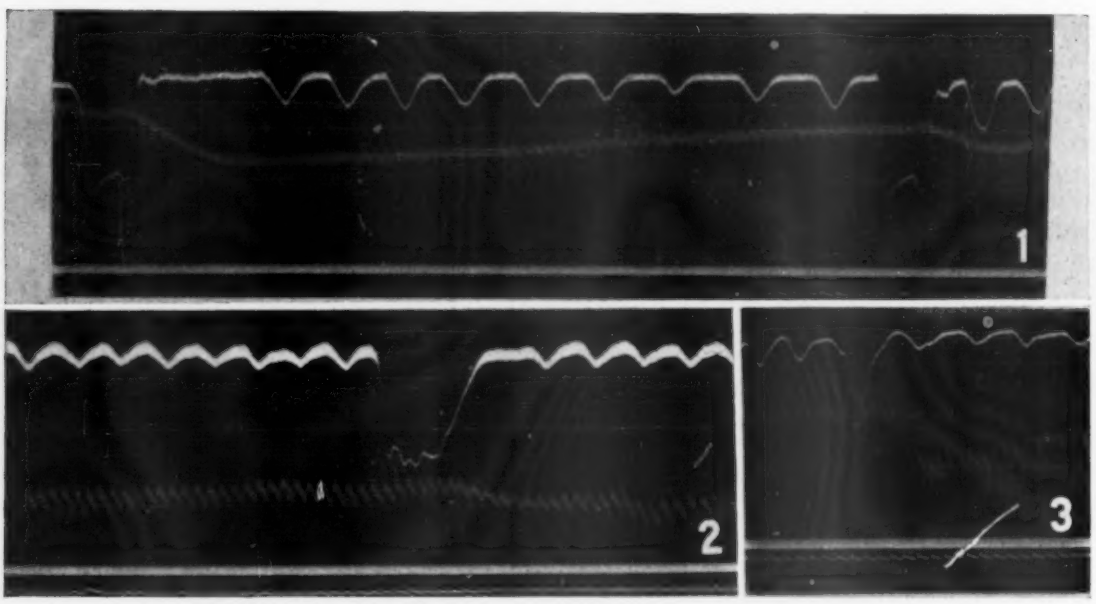

Fig. 1. Record of vasoconstriction in normal finger accompanying yawning.
Fig. 2. Record of vasoconstriction in sympathectomized finger accompanying yawning.
Fig. 3. Record of vasoconstriction in normal finger accompanying deep breath. The tracings from above downward are: (a) respiration by stethograph, (b) digital translucency, (c) signal magnet and (d) time in one second intervals.

completely formed (*Meroanenzephalie mit Meroankranie*). These observations suggest that if a yawn-center exists it is to be found in the medulla oblongata, possibly in the immediate vicinity of the respiratory and vasomotor centers.

The so-called psychically released yawning, or yawning as an apparent consequence of suggestion, implies that the act can be initiated by influences emanating from the highest cerebral levels. Inhibition of the act by influences arising at these levels is confirmed by the behaviour of hemiplegic patients who have lost the ability to fractionate the yawn-stretch act. Both Dumpert (24) and Lewy (35) have noted yawning in such patients attended by stretching movements confined to the affected (*i.e.*, "paralyzed") extremities. . . .doubtless to the amazement of the patient if not the doctor. Dumpert emphasizes that although this phenomenon is not to be seen in all hemiplegics it is consistently present in given cases.

Activation and possibly inhibitory modification of yawning by neural elements below the telencephalon is of course attested to by the observations of Gamper (46) and of Catel and Krauspe (47) on their cases of cerebral malformation. The involuntary nature of most yawns and their occasional occurrence in comatose patients (29) likewise argue for activation from so-called "sub-conscious levels," but nothing further is known about the afferent influences exciting this act.

Information regarding the centrally located efferent pathways subserving the yawn-stretch act is limited to deductions suggested by the involuntary activation of paralytic limbs during yawning. This stretching of limbs deprived of cortico-spinal control appears to indicate that the so-called "extra-pyramidal" pathways play a significant rôle in the uninhibited act.

Special sources. (Theories.) Numerous hypotheses concerning the mechanism and significance of yawning and stretching have appeared, and existing data do not suffice to show which of these are tenable. They are not mutually exclusive and it is possible that each contains a grain of truth.

The first theory to find its way into print, that of the Delsarteans, appeared in Russell's monograph (18) of 1891. This little book, which remains the most exhaustive account of yawning in the English language, defies description and should be consulted in the original by anyone interested in whimsical physiology. Suffice it here to say that yawning is presented as an "automatic impulse" caused by "bad air in the lungs," designed by Nature as a "gymnastic," and intended both to "awaken the respiratory organs into activity" and to "effect a stimulation of the brain through increased activity of the circulation." Although this concept enjoys no supporting evidence it appears to be the fore-runner of present day lay belief (16, 17) and it is not without current favor among clinicians. Within the past decade Otto (33) and Seeligmüller (48) have reiterated the notion that yawning is a gymnastic the practice of which should be cultivated as a means of toning up the body or relaxing mental tension.

Hauptmann (26) appears to have been the first clinician to attempt an interpretation of the yawn-stretch act. Writing in 1920, he suggested that this complex is a means of combating the loss of muscle tone which, along with tedium, develops as a consequence of enforced inactivity of the higher cerebral centers. He emphasized that this augmentation of muscle tone and metabolism, rather than the deep inspiratory effort, is the main purpose of the yawn-stretch act. We have already noted his contention that any temporary improvement in aeration due to this inspiratory effort is more than offset by the period of apnea which follows the expiratory phase. By this process of reasoning, but without recourse to precise measurements, Hauptmann rejects the notion that yawning bespeaks an anoxia of cerebral tissue. This conclusion has been widely accepted (24, 27, 29, 35) without confirmatory evidence.

Mayer (29) concluded that the yawn-stretch act is an automatic expression of cerebral fatigue in the sense that crying and laughing are expressions of sorrow and happiness. In this conclusion he has been joined by several subsequent writers (10, 27, 35). Mayer goes on to suggest that the act accomplishes im-

Crämer that he postulated an auto-intoxication by bacterial toxins arising from within the bowel as the causative agent. He reviews the rather meagre evidence supporting this view.

Neurologists have met with abnormal yawning rather more frequently than others. Oppenheim (37), Geigel (34) and Lewy (35) have encountered it in hysteria, and its occurrence as an aura in epilepsy has been remarked by Oppenheim (37), Lewy (35), Wilson (38) and Penfield (39). In Wilson's experience this aura has been associated with his so-called "peri-ventricular" or "visceral" epilepsy wherein the yawn is followed by voiding of urine, giddiness, headache and sometimes vomiting. He suggests that such cases depend on disorder of visceral centers in the third ventricle regulating sleep and fluid interchange, with subsequent spread to others on the floor of the fourth ventricle. Penfield (p. 110) gives the clinical details of such a case.

Among the manifold sequelae of epidemic encephalitis is that of paroxysmal yawning usually in conjunction with other, though varied, neurologic disturbances. Mayer (29), Wilson (38) and Sicard and Paraf (40) describe cases from their personal experience, but it has not been possible to ascribe the symptom to proven lesions at particular foci in the neuro-axis. In this connection it is of interest to record certain impressions which Yakovlev (41) has formed with regard to derangements of yawning and pandiculation in other disorders involving the basal ganglia. In his experience Huntington's Choreics exhibit a propensity for yawning and more especially for stretching which stands in sharp contrast to the behaviour of Parkinsonics who rarely yawn and never stretch. Yakovlev emphasizes that these impressions have been derived from a study of relatively few cases and stand in need of confirmation by a test on more material.

Excessive yawning is frequently seen in patients harboring expanding intracranial lesions, notably cerebellar abscess or tumor (35, 36, 37, 38, 42). When attended by abnormal drowsiness the sign is without localizing value (37, 38), but its occurrence in an alert patient suggests to some (38) that a thorough search be made for third ventricular and hypothalamic signs. Nash (36) hazards an explanation for the yawning seen in the usual run of brain tumor cases. He states:

> The occurrence of yawning, which is essentially an atypical respiratory act, in cases of brain tumor can be explained by irritation of the brain resulting in "afferent" impulses to the medullary respiratory center, which modify its rhythmic activity. (p. 92)

Special sources. (Investigative.) Mayer (29) alone has attempted a systematic investigation of the yawn-stretch act. By means of the roentgen-screen, laryngeal mirror and palpation of muscles he has analyzed the sequence of events in man when the complete complex, unmodified by inhibition, is allowed to develop. For convenience of description he divides the act into a number of phases:

I. Inspiratory Phase (duration: 4.4 to 6.8 sec.)
  A. Initial Part
  B. Acme (duration: 2.5 sec.)
II. Expiratory Phase

During the initial part of the inspiratory phase one observes a widening of the chest, a descent of the diaphragm and larynx, an elevation of the *ali nasi* and soft palate, a downward and backward displacement of the tongue, an abduction of the vocal cords and an opening of the mouth. This is occasionally attended by a subjective crackling in the ears which is thought (43) to be due to a contraction of the *M. tensor veli palatini* with jerking of its tendon in the bursa between it and the pterygoid hamulus.

When the acme of the complex has been attained one notes that the chest, diaphragm, larynx, tongue, palate and mouth have maintained or accentuated their respective movements of the initial phase. There has, in addition, been a firm closure of the eyes and a rapid development of the stretch component. The number and distribution of muscles recruited for the stretch are subject to variation and may be modified by voluntary inhibition. Neck and arm muscles are most frequently recruited, but those of the legs, trunk and abdomen may participate. The probable influence of such widespread muscular activity upon the venous return, and secondarily upon cardiac filling, rate and output, has been recognized (24, 29), but no studies have been made to confirm this. At the close of this phase saliva occasionally spurts from the submaxillary and sublingual ducts, presumably due to a squeezing of these glands by the contracting oral musculature.

During the expiratory phase one notes a short sharp expiration with gradual relaxation of all previously activated muscles permitting the return of diaphragm, larynx, tongue, *etc.* to their resting positions. Lacrimation and swallowing may occur at this stage. The former is believed (44) to be due to an augmentation of secretion rather than to a simple compression of the lacrymal glands because the tightest possible voluntary blepharospasm cannot squeeze forth tears in an amount sufficient to overflow the conjunctival sac. The swallowing is presumed to depend upon the occurrence of the previously noted salivation. This final phase of the yawn is usually accompanied by some such subjective feeling as relief, satisfaction or refreshment.

Prior to Mayer's (29) fluoroscopic observations, which show the diaphragm to descend and remain relatively fixed in this position from the initiation until the expiratory phase of the yawn, Hauptmann (26) had arrived at a different opinion as to the behaviour of this organ. Using a kymographic method for recording thoracic and abdominal respiration, Hauptmann was led to believe that the rhythmic action of this muscle was but little, if at all, interrupted by yawning. His view must, of course, give way to that suggested by the more refined method of fluoroscopy. It is important to note, however, that Mayer and Hauptmann agree that yawning effects no ultimate increase in aeration of the lungs; they argue that the apneic period following the yawn (see fig. 1 and 2) more than compensates for any temporary increase in aeration afforded by a single deep breath.

The neural structures subserving the complicated yawn-stretch act may conveniently be thought of as comprising an integrative center played upon by diverse afferent elements and discharging over its several efferent channels. Evi-

dence at hand is insufficient, however, to establish the location or even the existence of a "yawn-center" in man or animals. Ellenberger and Scheunert (45) include yawning among the functions of the medulla oblongata in dogs, but they do not submit the evidence supporting this localization. In man a postulated center has been variously located in the *corpus striatum* (23, 35) and "subcortical ganglia" (24, 27, 29) on very unsatisfactory evidence. There are, however, two significant papers bearing on this question. Gamper (46) has seen the complete yawn-stretch act in what he describes as a mid-brain preparation (*Arhinencephalie mit Encephalocele*), and Catel and Krauspe (47) have observed it in a case where only the medulla oblongata and more caudal portions of the nervous system were

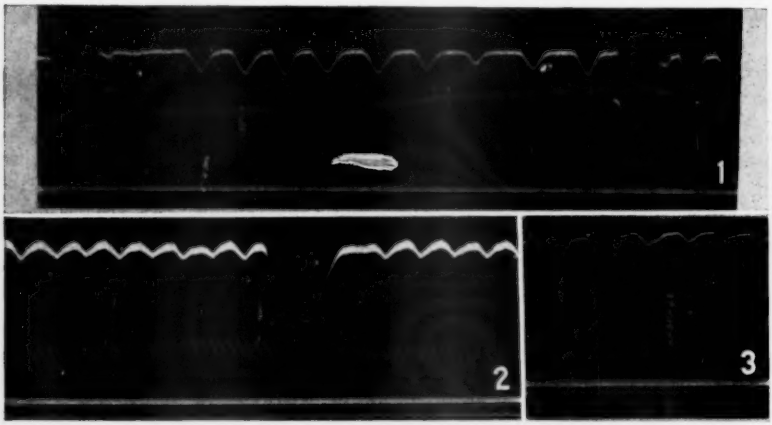

Fig. 1. Record of vasoconstriction in normal finger accompanying yawning.
Fig. 2. Record of vasoconstriction in sympathectomized finger accompanying yawning.
Fig. 3. Record of vasoconstriction in normal finger accompanying deep breath. The tracings from above downward are: (a) respiration by stethograph, (b) digital translucency, (c) signal magnet and (d) time in one second intervals.

completely formed (*Meroanenzephalie mit Meroankranie*). These observations suggest that if a yawn-center exists it is to be found in the medulla oblongata, possibly in the immediate vicinity of the respiratory and vasomotor centers.

The so-called psychically released yawning, or yawning as an apparent consequence of suggestion, implies that the act can be initiated by influences emanating from the highest cerebral levels. Inhibition of the act by influences arising at these levels is confirmed by the behaviour of hemiplegic patients who have lost the ability to fractionate the yawn-stretch act. Both Dumpert (24) and Lewy (35) have noted yawning in such patients attended by stretching movements confined to the affected (*i.e.*, "paralyzed") extremities. . . .doubtless to the amazement of the patient if not the doctor. Dumpert emphasizes that although this phenomenon is not to be seen in all hemiplegics it is consistently present in given cases.

Activation and possibly inhibitory modification of yawning by neural elements below the telencephalon is of course attested to by the observations of Gamper (46) and of Catel and Krauspe (47) on their cases of cerebral malformation. The involuntary nature of most yawns and their occasional occurrence in comatose patients (29) likewise argue for activation from so-called "sub-conscious levels," but nothing further is known about the afferent influences exciting this act.

Information regarding the centrally located efferent pathways subserving the yawn-stretch act is limited to deductions suggested by the involuntary activation of paralytic limbs during yawning. This stretching of limbs deprived of corticospinal control appears to indicate that the so-called "extra-pyramidal" pathways play a significant rôle in the uninhibited act.

Special sources. (Theories.) Numerous hypotheses concerning the mechanism and significance of yawning and stretching have appeared, and existing data do not suffice to show which of these are tenable. They are not mutually exclusive and it is possible that each contains a grain of truth.

The first theory to find its way into print, that of the Delsarteans, appeared in Russell's monograph (18) of 1891. This little book, which remains the most exhaustive account of yawning in the English language, defies description and should be consulted in the original by anyone interested in whimsical physiology. Suffice it here to say that yawning is presented as an "automatic impulse" caused by "bad air in the lungs," designed by Nature as a "gymnastic," and intended both to "awaken the respiratory organs into activity" and to "effect a stimulation of the brain through increased activity of the circulation." Although this concept enjoys no supporting evidence it appears to be the fore-runner of present day lay belief (16, 17) and it is not without current favor among clinicians. Within the past decade Otto (33) and Seeligmüller (48) have reiterated the notion that yawning is a gymnastic the practice of which should be cultivated as a means of toning up the body or relaxing mental tension.

Hauptmann (26) appears to have been the first clinician to attempt an interpretation of the yawn-stretch act. Writing in 1920, he suggested that this complex is a means of combating the loss of muscle tone which, along with tedium, develops as a consequence of enforced inactivity of the higher cerebral centers. He emphasized that this augmentation of muscle tone and metabolism, rather than the deep inspiratory effort, is the main purpose of the yawn-stretch act. We have already noted his contention that any temporary improvement in aeration due to this inspiratory effort is more than offset by the period of apnea which follows the expiratory phase. By this process of reasoning, but without recourse to precise measurements, Hauptmann rejects the notion that yawning bespeaks an anoxia of cerebral tissue. This conclusion has been widely accepted (24, 27, 29, 35) without confirmatory evidence.

Mayer (29) concluded that the yawn-stretch act is an automatic expression of cerebral fatigue in the sense that crying and laughing are expressions of sorrow and happiness. In this conclusion he has been joined by several subsequent writers (10, 27, 35). Mayer goes on to suggest that the act accomplishes im-

portant but as yet undefined physiologic adjustments in the state of the circulation and in the metabolism of the recruited muscles.

Of all hypotheses advanced that of Dumpert (24) is by far the most elaborate. The salient features of his concept are as follows:

1. Yawning (the mouth act attended by a deep inspiration) is but part of the larger yawn-stretch complex which occurs in many animals, in infants and in children. This larger reaction is first regularly dissociated into its respiratory and stretching components in adult man, and even here the ability to fractionate the act can disappear following certain lesions of the central nervous system.
2. The yawn-stretch act is believed to occur whenever the cerebral circulation is not favorably adjusted for maintenance of the waking and alert states.
3. The yawn-stretch complex or even yawning alone is thought to effect an augmentation of the venous return to the heart by means of:
   a. The "milking action" of the striated muscles upon the veins of the limbs;
   b. The compressing action of the contracting abdominal muscles and the descending diaphragm upon the veins of the splanchnic bed; and,
   c. The aspirating effect of the lowered intra-thoracic pressure upon blood entering the *venae cavae* during the inspiratory phase of the act.
4. This presumed increase in venous return is thought to effect an increase in the rates of cardiac filling and output (Bainbridge's reflex), thereby providing a generalized improvement in the supply of arterial blood to the tissues.
5. A differential distribution of the supposedly augmented arterial flow is presumed to occur for the especial benefit of the cerebral circulation. This selective redistribution is believed traceable to the effects of centripetal impulses arising in the stretched muscles and serving to:
   a. Produce a widespread systemic vasoconstriction by their action on the vasomotor center (a la Bayliss); and,
   b. Effect a cerebral vasodilatation by their action on Weber's center.
6. Thus the yawn-stretch reaction and, less effectively, the yawn alone are able to combat cerebral circulatory states unfavorable to the maintenance of the waking and alert conditions.

Although much that Dumpert postulates is probable and my own observations (*vide infra*) confirm the occurrence of the systemic vasoconstriction he predicted, his theory is without support at its crucial points. Thus it is unjustifiable to assume that the cerebral circulation is diminished during sleep; indeed, the only reliable information (49) we have on the state of the cerebral circulation prior to, during and after sleep fails to disclose any significant variations. Further, there are no observations confirming an increase of the venous return during yawning. Finally, Dumpert invokes the aid of a cerebro-dilatory center the existence of which is open to question.

Peiper (27) looks upon yawning as a purely respiratory act which represents a transient disorder of the respiratory center. He believes that as a result of fatigue the excitability or sensitivity (*erregbarkeit*) of the higher respiratory centers (50) is depressed whereupon a lower lying yawn-center is released. He thus explains the occurrence of an automatic expression of fatigue and does not postulate that it accomplishes any physiologic adjustments. His theory is attractive but rests upon purely hypothetical considerations.

One further theory concerning the yawn has come to my attention. It has been suggested that as the nuchal muscles are recruited in the stretch they squeeze

upon the thyroid gland, express thyroxine into the blood stream and thereby accelerate the metabolism. The literature has been searched in vain for the origin and subsequent reference to this idea.

NEW OBSERVATIONS. *Methods and material.* My observations on yawning were made while studying the vasomotor activity in the digits of adult humans some of whom were "normal" and others of whom were victims of either Raynaud's or (early) Buerger's diseases. The diseased group was studied both before and after therapeutic pre-ganglionic sympathectomy. None was aware of my interest in yawning.

The subject, comfortably seated, placed one finger or toe, without restraining device, on a simple rest between a photo-electric cell and a constant light source. Current set up in the cell by light passing through the digit was led to a 2-stage amplifier, whence it emerged to activate a DuBois oscillograph. A mirror mounted on the latter directed a beam of reflected light onto a moving strip of bromide paper. The electrical circuit employed was such that a downward deflection on the photographic record was produced by an increase in the amount of light (and decrease in the amount of blood) traversing the digit.

A stethograph of the air-transmission type was applied to the chest and its indicator made to cast a second shadow on the sensitized paper. A downward excursion of this indicator was produced by an inspiratory movement.

Tracings of a time-marker and a signal magnet also appear on the photographic records.

Subjects were routinely observed over periods of 30 to 60 minutes in a warm darkened room. Under these conditions spontaneous yawning was frequent.

Observations. In the digits of normal adults a vasoconstriction was regularly observed to follow a spontaneous yawn. Figure 1 is a characteristic record of the development and subsidence of this phenomenon in a finger. From this tracing it can be seen that the vasoconstriction begins to appear after a latency of about 4 to 4.5 seconds as measured from the inception of the inspiratory phase of the yawn. The response is maximal within 9 to 10 seconds after the onset of the yawn and then gradually subsides during the ensuing 45 seconds. In a large number of such observations the latent period was regularly found to be between 3.5 and 5 seconds. The magnitude of the response and the time required for its complete development and subsidence were subject to considerable variation even in the same individual. In general these values were roughly proportional to the depth and duration of inspiratory movement. Maximal response was nearly always evident within 7 to 12 seconds and subsidence was usually complete within 30 to 60 seconds.

It was also observed that yawning in normal adults is regularly attended by a transient cardiac acceleration. This speeding of the heart was found to develop and subside *pari passu* with the vasoconstrictor response. The subject in figure 1, whose resting pulse is 80/min., shows an acceleration to 90/min. This degree of cardiac acceleration is fairly representative of many similar observations but it is subject to variation among different individuals and even in the same individual according to the depth of the yawn.

That the observed changes in the translucency of digits during a yawn are truly vasoconstrictor responses dependent upon the integrity of the autonomic innervation of the part is clearly shown by studies on sympathectomized extremities. Prior to surgical intervention the digits of such limbs exhibited responses equal to or even greater than those observed in normal parts. Following sympathectomy they show a greatly diminished or no vasomotor response depending upon the completeness of the denervation. Figure 2, a tracing from the finger of a Raynaud patient after sympathectomy, shows no appreciable vasoconstriction following a deep yawn. It is interesting to note, however, that the cardiac acceleration is not abolished. This subject's resting pulse rate of 70/min. is raised to one of 80 to 85/min. during the inspiratory phase of the yawn.

Results in every way comparable to those depicted for fingers were obtained from the toes of normal and sympathectomized subjects.

Discussion. Although Dumpert (24) and Mayer (29) predicted that circulatory alterations would be found to accompany yawning, there has been, to my knowledge, no previous demonstration of such changes. The phenomena under consideration are, however, closely related to or identical with the circulatory changes which Bolton *et al.* (51) describe as following a deep inspiration.

In a series of experiments on man Bolton *et al.* showed that following a deep voluntary inspiration a vasoconstriction occurs in the digital vessels. Its reflex nature was conclusively established by experiments on denervated and sympathectomized limbs, and some evidence was adduced in favor of the afferent stimulus arising from the expansion of the chest wall. Figure 3 is a recording by our apparatus of the vasomotor reflex they describe.

In the discussion of their observations Bolton *et al.* make no mention of their reflex being accompanied by a cardiac acceleration. Careful scrutiny of their figures 2 and 3 suggests, however, that this did occur in their experiments. The subject of my figure 3 exhibited a resting pulse rate of 70/min. This rose to 80/min. during the 10 seconds following the initiation of the deep inspiration and is typical of the behaviour of the pulse in many such observations.

SUMMARY

The literature on yawning and stretching has been reviewed. The descriptive phase of the subject is essentially complete for man. The experimental phase is barely begun with the demonstration that circulatory changes consisting of transient cardiac acceleration and digital vasoconstriction attend spontaneous yawning in man. Aspects of the problem remaining completely unexplored include:

a. The behaviour of the cerebral circulation;
b. Measurements of cardiac filling and output;
c. The chemistry of the respiratory component;
d. The nature of the motor discharge in the stretch component; and,
e. The possibility of endocrine (*e.g.*, thyroid) changes.

Until such observations are completed, any formulation of the physiological significance of this act will rest upon an insecure foundation.

REFERENCES

(1) Finesinger, J. E., A. P. Heusner and R. H. Smithwick. Vasomotor studies in Raynaud's disease before and after sympathectomy. Tr. Am. Neurol. A. **65:** 193, 1939.

(2) Bayliss, W. M. Principles of general physiology, 2nd ed. Revised, Longmans, Green & Co., London, 1918.

(3) Lehrbuch der Allgemeinen Physiologie. Ed. by E. Gellhorn; Thieme, Leipzig, 1931.

(4) Wright, S. Applied physiology. 5th ed., Oxford Univ. Press, New York, 1934.

(5) Winton, F. R. and L. E. Bayliss. Human physiology. 2nd ed., Blakiston's Son & Co., Inc., Philadelphia, 1937.

(6) Macleod's Physiology in modern medicine. Ed. by P. Bard; 8th ed., Mosby Co., St. Louis, 1938.

(7) Howell, W. H. A textbook of physiology. 14th ed., Saunders Co., Philadelphia, 1940.

(8) Starling's principles of human physiology. Ed. by C. L. Evans; 8th ed., Lea & Febiger, Philadelphia, 1941.

(9) Best, C. H. and N. B. Taylor. The physiological basis of medical practice. 3rd ed., Williams & Wilkins, Baltimore, 1943.

(10) Wiggers, C. J. Physiology in health and disease. 3rd ed., Lea & Febiger, Philadelphia, 1939, p. 386.

(11) Starling, E. H. Text-book of physiology. Ed. by E. A. Schäfer; Pentland, Edinburg & London, 1900, Vol. ii, p. 309.

(12) Rosenthal, J. Handbuch der Physiologie. Ed. by L. Hermann; Vogel, Leipzig, 1880, B. IV, H. 2, S. 234.

(13) Boruttau, H. Handbuch der Physiologie des Menschen. Ed. by W. Nagel; Vieweg u. Sohn, Braunschweig, 1905, B. I, S. 27.

(14) Dictionary of psychology. Ed. by H. C. Warren; Houghton Mifflin Co., New York, 1934.

(15) Moore, J. E. Some psychological aspects of yawning. J. Gen. Psychol. **27:** 289, 1942.

(16) New York Times Magazine, October 29, 1944.

(17) Clarkson, H. E., Jr. When the yawn breaks. Hygeia **20:** 585, 1942.

(18) Russell, H. Yawning. The Delsarte Series, No. 1, U. S. Book Company, New York, 1891.

(19) Shorter Oxford English dictionary. 2nd ed., Clarendon Press, Oxford, 1939.

(20) Webster's New International Dictionary. 2nd ed., Unabridged, G. & C. Merriam Co., Springfield, 1942.

(21) Gould's Medical Dictionary. 4th ed., by G. M. Gould; Blakiston, Philadelphia, 1935.

(22) The American Illustrated Medical Dictionary. 20th ed., by W. A. N. Dorland; Saunders Co., Philadelphia, 1944.

(23) Crämer, F. Über Sodbrennen und Gähnen. Arch. f. Verdauungskr. **33:** 149, 1924.

(24) Dumpert, V. Zur Kenntnis des Wesens und der physiologischen Bedeutung des Gähnens, J. f. Psychol. u. Neurol. **27:** 82, 1921.

(25) Heinroth. Quoted by Dumpert (24).

(26) Hauptmann, A. Wie, wann und wozu gähnen wir? Neurol. Zentralbl. **39:** 781, 1920.

(27) Peiper, A. Das Gähnen. Deutsch med. Wchnschr. **58:** 693, 1932.

(28) Darwin, C. The expression of the emotions in man and animals. Appleton, New York, 1873.

(29) Mayer, C. Physiologisches und Pathologisches über das Gähnen. Ztschr. f. Biol. **73:** 101, 1921.

(30) Schrijver, J. Über das Magengähnen. Arch. f. Verdauungskr. **43:** 378, 1928.

(31) Luciani. Physiologie des Menschen. T. 1, Jena, 1911; quoted by Schrijver (30).

(32) Carlson, A. J. Personal communication to Schrijver (30).

(33) Otto, H. Vom Recken, Strecken, Gähnen und Husten. Fortschr. d. Med. **53:** 304, 1935.

(34) GEIGEL, R. Das Gähnen. Muenchen med. Wchnschr. **55**: 223, 1908.
(35) LEWY, E. Über das Gähnen. Ztschr. f. Neurol. u. Psychiat. **72**: 161, 1921.
(36) NASH, J. Surgical physiology. Thomas, Springfield, 1942, pp. 92 and 115.
(37) OPPENHEIM, H. Lehrbuch der Nervenkrankheiten. 5th ed., Karger, Berlin, 1908, pp. 1219 and 1445.
(38) WILSON, S. A. K. Text book of neurology. Ed. by A. N. Bruce; Williams & Wilkins, Baltimore, 1940; Vol. i, p. 121, and Vol. ii, pp. 1223 and 1507.
(39) PENFIELD, W. AND T. C. ERICKSON. Epilepsy and cerebral localization. Thomas, Springfield, 1941, p. 110.
(40) SICARD, J. A. ET J. PARAF. Fou Rire Syncopal et Baillments au Cours de L'Encephalite Epidémique. Bull. et mem. de soc. méd. d. hôp. de Paris **45**: 232, 1921.
(41) YAKOVLEV, P. I. Personal communication.
(42) MACEWEN, W. Pyogenic infective diseases of the brain and spinal cord. Maclehose & Sons, Glasgow, 1893, p. 195.
(43) VALENTIN. Ztschr. f. Ohrenheilk. **46**: 84; quoted by MAYER (29).
(44) RÖMER. Lehrbuch der Augenheilkunde. 3 aufl., 1919, p. 241; quoted by MAYER (29).
(45) ELLENBERGER, W. UND A. SCHEUNERT. Lehrbuch der vergleichenden Physiologie der Haussäugetiere. Parey, Berlin, 1910, p. 578.
(46) GAMPER, E. Bau und Leistungen eines menschlichen Mittelhirnwesens (Arhinencephalie mit Encephalocele) zugleich ein Beitrag zur Teratologie und Fasersystematik. Ztschr. f. d. ges. Neurol. u. Psychiat. **104**: 49, 1926.
(47) CATEL, W. UND C. A. KRAUSPE. Über die nervöse Leistung und den anatomischen Bau einer menschlichen Hirnmissbildung (Meroanenzephalie mit Meroakranie). Jahrb. f. Kinderh. **129**: 1, 1930.
(48) SEELIGMÜLLER, A. Von Gähnen und Räkeln. Therap. d. Gegenw. **81**: 440, 1940.
(49) GIBBS, F. A., E. L. GIBBS AND W. G. LENNOX. The cerebral blood flow during sleep in man. Brain **58**: 44, 1935.
(50) LUMSDEN, T. Observations on the respiratory centres in the cat. J. Physiol. **57**: 153, 1923.
(51) BOLTON, B., E. A. CARMICHAEL AND G. STÜRUP. Vaso-constriction following deep inspiration. J. Physiol. **86**: 83, 1936.